The international journal of science / 8 June 2023

outlook
Smoking

nature



GROWTH FACTOR

How warmer waters helped guide the evolution of small, fast-growing coral-reef fishes

Lofty ambitions
The epic quest to reach record-breaking trees in the Amazon

Software engineered
Game-playing AI speeds up sorting in computer code

Crown jewellery
DNA reveals details of who held Stone Age tooth pendant

VISIT…

LANZAROTE Caliente.com

Nature.2023.06.10

[Sat, 10 Jun 2023]

- **This Week**
- **News in Focus**
- **Books & Arts**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**
- **Nature Outlook**

This Week

- **The inspiring story of the Tara and its 20-year message from the corals** [06 Jun 2023]
  • A remarkable research ship's voyages to understand marine biodiversity show how visionary thinking can boost understanding of the natural world — and help to better preserve it.
- **Why Nature will not allow the use of generative AI in images and video** [07 Jun 2023]
  • Saying 'no' to this kind of visual content is a question of research integrity, consent, privacy and intellectual-property protection.
- **The US Supreme Court has gutted federal protection for wetlands — now what?** [05 Jun 2023]
  • Most US wetlands just lost federal protection, but there's still time for state and local governments to act. Scientists and the public can help.

- EDITORIAL
- 06 June 2023

The inspiring story of the *Tara* and its 20-year message from the corals

A remarkable research ship's voyages to understand marine biodiversity show how visionary thinking can boost understanding of the natural world — and help to better preserve it.



The *Tara* during its 100,000-kilometre voyage around the Pacific Ocean.Credit: François Aurat/Tara Ocean Foundation

Scientists have known for decades that the biodiversity we know is a fraction of the biodiversity that exists. Even when it comes to perhaps the simplest measure, the number of species, researchers estimate that there could be anywhere between 3 million and 100 million species, of which some 1.7 million have been described. Every year, the names of some 10,000 new ones get added to the list.

Gathering good biodiversity data can be a mission in itself, especially on marine biodiversity. One dependable source is a schooner called *Tara*, which celebrates 20 years at sea as a research ship this year. *Tara* has been to the Arctic Ocean and the Mediterranean Sea. Last week, researchers reported the results of its latest voyage, Tara Pacific, a two-year expedition across the Pacific Ocean, published in a collection of articles in Springer Nature journals (go.nature.com/45puzhk).



Epic voyage finds astonishing microbial diversity among coral reefs

Coral reefs are among Earth's most diverse ecosystems, supporting 25% of marine life and providing services such as food, jobs and coastal protection to nearly one billion people worldwide. One region of the Pacific, the Coral Triangle — which includes the waters of Indonesia, Malaysia and the Philippines — has around 75% of the world's coral species. But these nurseries for marine life are under threat: globally, around 50% of living corals have been lost since the 1950s (T. D. Eddy *et al. One Earth* **4**, 1278–

1285; 2021). Moreover, climate change poses a large risk to their continued survival.

One of Tara Pacific's research groups has been focusing on genetic diversity — in particular, that of microbial communities (bacteria and archaea) living in Pacific Ocean corals. The scientists, led by marine microbiologist Pierre Galand at the Banyuls Oceanological Observatory in Banyuls-sur-Mer, France, collected more than 5,000 samples, and focused on analysing the 16S ribosomal DNA marker gene, which is used to identify and classify microorganisms. They found around half a million distinct DNA sequences known as amplicon sequence variants (ASVs), which can be used as a measure of the genetic diversity in a sample (*Nature* https://doi.org/kddz; 2023). From these data, the researchers estimated that the microbial diversity of coral reefs globally is probably around 2.8 million ASVs (P. E. Galand *et al. Nature Commun.* **14**, 3039; 2023). For comparison, this is close to the lower end of one genetic-diversity estimate of all of Earth's bacterial and archaeal communities — a proposed range of 2.72 million and 5.44 million ASVs (S. Louca *et al. PLoS Biol.* **17**, e3000106; 2019). Galand and his colleagues' work, which builds on smaller-scale studies of coral reefs (M. Chiarello *et al. Proc. R. Soc. B* **287**, 20200642; 2020), confirms that Earth's microbial genetic diversity is much higher and richer than previously thought.



Collection: Tara Pacific

The research ship has a stirring and unusual back story. Its original captain was Peter Blake, a much-decorated professional yachtsman from New Zealand. After retirement, Blake became an environmental envoy to the United Nations but was killed by pirates at the mouth of the Amazon River while on an expedition in 2001. Agnès Troublé, a French fashion designer known as Agnès B, and her son Etienne Bourgois acquired the boat, determined to continue Blake's original vision. They established the Tara Ocean Foundation and invited scientists and research funders to join them on various missions.

The logistics of converting a schooner into a floating laboratory and taking it on a 100,000-kilometre journey cannot be underestimated, especially considering the present complicated relations between Pacific nations. And then there's the logistics of the research itself: organizing 3,000 dives; sending samples for PCR analysis en route; keeping the voyage on track.

Researchers must continue to build on the work being reported and refine our understanding of the importance of diversity for safeguarding ecosystem stability and function. The project is a great example of visionary thinkers such as Troublé and Bourgois working closely with funders and scientists to help us to understand the breathtaking diversity of the world around us before it is too late. When it comes to preserving the natural systems we all depend on, more such collaborations are needed.

Nature **618**, 213 (2023)

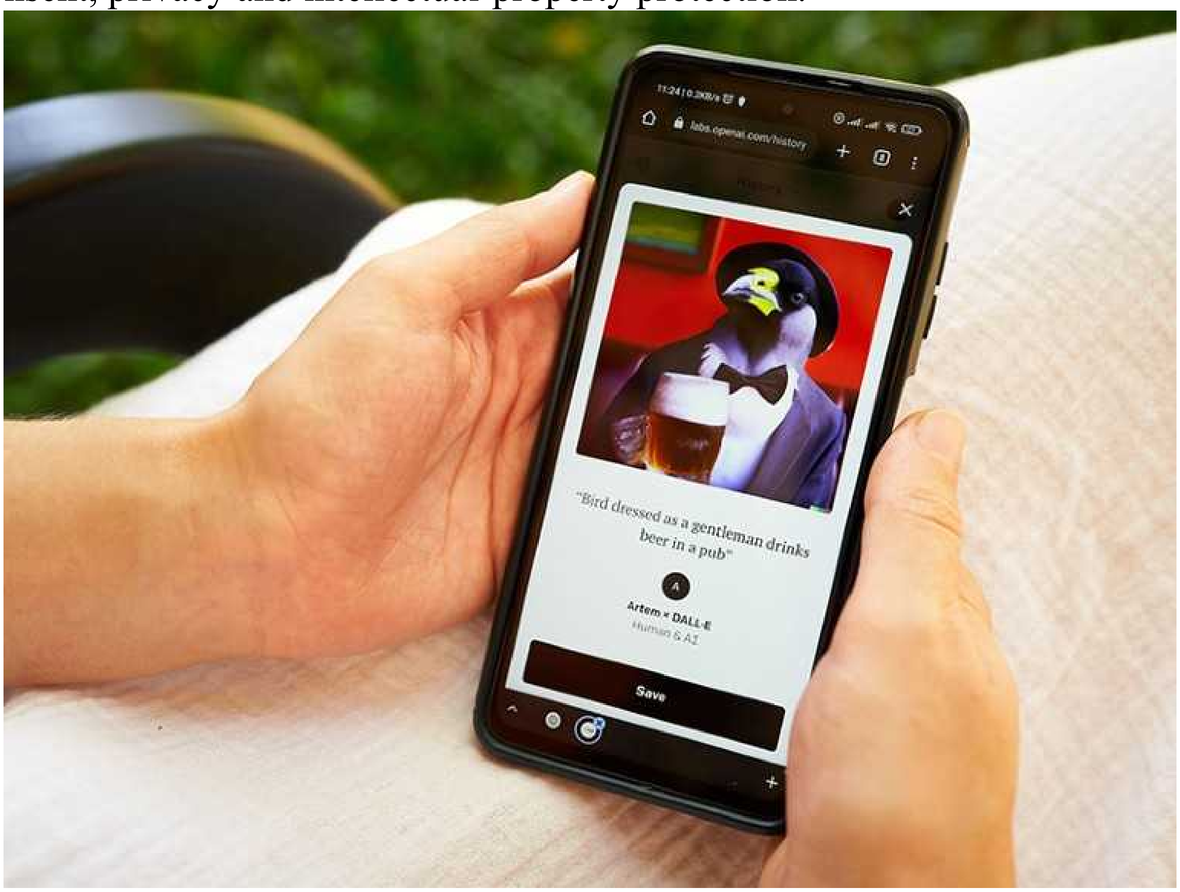
doi: https://doi.org/10.1038/d41586-023-01826-z

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01826-z

- EDITORIAL
- 07 June 2023

Why *Nature* will not allow the use of generative AI in images and video

Saying ‘no’ to this kind of visual content is a question of research integrity, consent, privacy and intellectual-property protection.

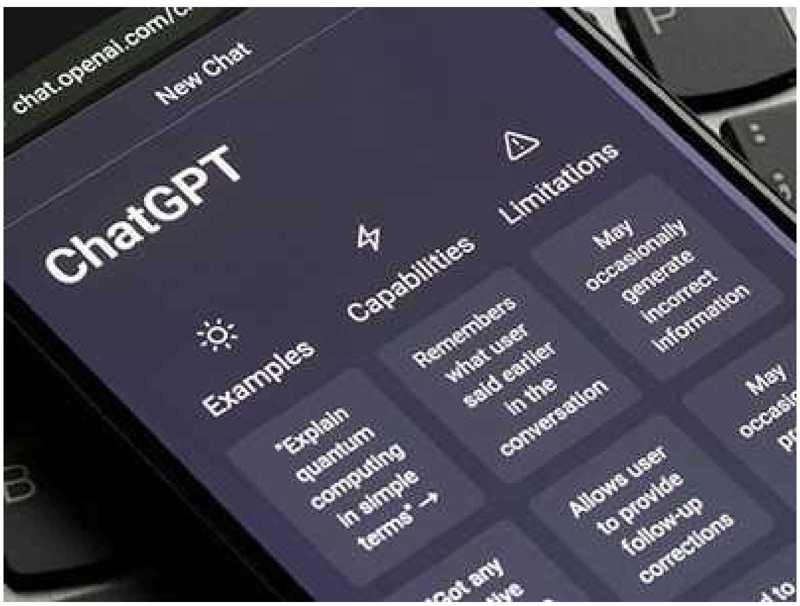


Nature will not publish imagery created wholly or partly using generative AI.Credit: Artem Medvediev/Alamy

Should *Nature* allow generative artificial intelligence (AI) to be used in the creation of images and videos? This journal has been discussing, debating and consulting on this question for several months following the explosion of content created using generative AI tools such as ChatGPT and Midjourney, and the rapid increase in these platforms' capabilities.

Apart from in articles that are specifically about AI, *Nature* will not be publishing any content in which photography, videos or illustrations have been created wholly or partly using generative AI, at least for the foreseeable future.

Artists, filmmakers, illustrators and photographers whom we commission and work with will be asked to confirm that none of the work they submit has been generated or augmented using generative AI (see go.nature.com/3c5vrtm).



Tools such as ChatGPT threaten transparent science; here are our ground rules for their use

Why are we disallowing the use of generative AI in visual content? Ultimately, it is a question of integrity. The process of publishing — as far as both science and art are concerned — is underpinned by a shared commitment to integrity. That includes transparency. As researchers, editors and publishers, we all need to know the sources of data and images, so that

these can be verified as accurate and true. Existing generative AI tools do not provide access to their sources so that such verification can happen.

Then there's attribution: when existing work is used or cited, it must be attributed. This is a core principle of science and art, and generative AI tools do not conform to this expectation.

Consent and permission are also factors. These must be obtained if, for example, people are being identified or the intellectual property of artists and illustrators is involved. Again, common applications of generative AI fail these tests.

Generative AI systems are being trained on images for which no efforts have been made to identify the source. Copyright-protected works are routinely being used to train generative AI without appropriate permissions. In some cases, privacy is also being violated — for example, when generative AI systems create what look like photographs or videos of people without their consent. In addition to privacy concerns, the ease with which these 'deepfakes' can be created is accelerating the spread of false information.

Appropriate caveats

For now, *Nature* is allowing the inclusion of text that has been produced with the assistance of generative AI, providing this is done with appropriate caveats (see go.nature.com/3cbrjbb). The use of such large language model (LLM) tools needs to be documented in a paper's methods or acknowledgements section, and we expect authors to provide sources for all data, including those generated with the assistance of AI. Furthermore, no LLM tool will be accepted as an author on a research paper.

The world is on the brink of an AI revolution. This revolution holds great promise, but AI — and particularly generative AI — is also rapidly upending long-established conventions in science, art, publishing and more. These conventions have, in some cases, taken centuries to develop, but the result is a system that protects integrity in science and protects content creators from exploitation. If we're not careful in our handling of AI, all of these gains are at risk of unravelling.

Many national regulatory and legal systems are still formulating their responses to the rise of generative AI. Until they catch up, as a publisher of research and creative works, *Nature*'s stance will remain a simple 'no' to the inclusion of visual content created using generative AI.

Nature **618**, 214 (2023)

doi: https://doi.org/10.1038/d41586-023-01546-4

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01546-4

| Section menu | Main menu |

- WORLD VIEW
- 05 June 2023

The US Supreme Court has gutted federal protection for wetlands — now what?



Most US wetlands just lost federal protection, but there’s still time for state and local governments to act. Scientists and the public can help.

- Royal C. Gardner

Each year, I take my environmental-law students on a swamp walk in Big Cypress National Preserve in the western Everglades in Florida. Hip-deep in the translucent water, and mindful of alligators, we appreciate the cathedral-like setting inside a cypress dome. The trip helps the students to understand aquatic ecosystems and how wetlands influence downstream waters. Wetlands provide a range of ecosystem services related to water quality, flood control, aquifer recharge, shoreline stabilization and species

habitats. A broader lesson is that environmental law, to be effective, must take science into account.

Sadly, the US Supreme Court's recent ruling in *Sackett* v. *Environmental Protection Agency* demonstrates that a majority of the justices never learnt that lesson. The case has eviscerated federal protection of wetlands under the Clean Water Act, the principal federal law responsible for the improvement of the country's water quality over the past 50 years. Before it, for example, oil and debris in the Cuyahoga River in Ohio frequently caught fire.

The court's decision is dire, but state and local governments still have the power to defend wetlands. There has never been a more important time for scientists and the public to champion state and local wetland protection, wherever they live.



The extent and drivers of global wetland loss

Signed in 1972, the Clean Water Act has prohibited discharging pollutants into "waters of the United States" or filling them in, unless a permit is granted. The issue in the *Sackett* case was whether, and to what extent, wetlands are "waters of the United States" and thus protected.

The court quashed clean-water protections twice. First, it declared that the Clean Water Act applies to wetlands only when they have a continuous surface connection to a permanent body of water, such as an ocean, river, stream or lake. Wetlands connected to other bodies through groundwater are excluded, as are seasonal wetlands, such as vernal pools and playa lakes. By some estimates, this requirement wipes out Clean Water Act coverage for more than half of the nation's wetlands. The cypress dome that my students and I visit dries up from March to May, except for some isolated holes dug by alligators. The Clean Water Act no longer protects it.

Second, and worse, the court decreed that, to qualify as a water of the United States, a wetland must be so inseparably bound to an ocean, river, stream or lake that it is "difficult to determine where the 'water' ends and the 'wetland' begins".

This requirement has no basis in law or science. As my colleagues and I explained in a brief submitted to the court on behalf of 12 scientific societies, wetland researchers can almost always determine where a body of water ends and a wetland begins. Typically, wetland vegetation begins at a certain water depth, easily visible to wetland scientists and consultants. It is hard to identify any wetland that could pass this requirement, other than some tidal wetlands and perhaps wetlands within the high-water mark of a river. Most of the wetlands in Everglades National Park — designated by the United States as being of international importance under the Ramsar Convention, the 1971 international wetland-protection treaty — can be delineated, and thus are no longer protected by the Clean Water Act.



Why we need a new economics of water as a common good

This means that a federal permit is no longer needed to fill in most US wetlands. Projects that were denied because of their expected negative environmental consequences will now be able to proceed. Because there is no need for a permit, there is now no federal requirement to offset the adverse impacts of the loss of these aquatic resources.

With each wetland that is filled, people's water quality will suffer, municipalities' risks of floods will increase and habitat for wetland-dependent species will be eliminated.

The court ignored not just science, but also the intent of Congress — which specifically included wetlands adjacent to larger water bodies when it amended the Clean Water Act in 1977 — and the policies of presidential administrations from Gerald Ford to Joe Biden. Even the most restrictive view of the geographic scope of the Clean Water Act, put forth by Donald Trump's administration, was too broad for this court.

Congress could correct the error through legislation that more clearly defines which wetlands are federally protected. Given the current political divisions, however, it is unlikely that Congress will act quickly, if at all.

Wetland protection is therefore in the hands of state and local governments. Although some states, such as California and Maryland, have their own robust wetland-protection laws, most — including large, populous ones such as Colorado, Georgia and Texas — do not. Some states have tethered their regulatory programmes to the federal programme — so, when federal jurisdiction shrinks, so will state jurisdiction.

Now is a crucial time for scientists and the public to educate state and local decision makers about how acutely we need wetlands. Scientific societies and organizations such as Engineers and Scientists Acting Locally can identify opportunities to influence policy. I hope that many scientists (and others with an appreciation of science) will seek appointment to zoning boards or run for state, county or local political office, which groups such as 3.14 Action, based in Washington DC, train scientists to do. Otherwise, developers will be free to destroy wetlands without heeding the benefits they provide to people and nature.

Nature **618**, 215 (2023)

doi: https://doi.org/10.1038/d41586-023-01827-y

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01827-y

News in Focus

- **Turkey's researchers fear loss of freedom after Erdoğan re-elected** [30 May 2023]
  • Researchers expect the administration to further restrict autonomy and free speech. Some say they will move away or retire.
- **US will vaccinate birds against avian flu for first time — what researchers think** [26 May 2023]
  • The country plans to vaccinate endangered condors in an effort to curb unprecedented H5N1 outbreaks.
- **Chronic stress can inflame the gut — now scientists know why** [25 May 2023]
  • Signals originating in the brain make their way to gut nerve cells, leading to a release of inflammatory chemicals.
- **AI intensifies fight against 'paper mills' that churn out fake research** [31 May 2023]
  • Text- and image-generating tools present a new hurdle for efforts to tackle the growing number of fake papers making their way into the academic literature.
- **The human brain's characteristic wrinkles help to drive how it works** [31 May 2023]
  • A model of the brain's geometry better explains neuronal activity than a model based on the 'connectome'.
- **China is using satellites to police the protection of nature — but will it work?** [26 May 2023]
  • Scientists say the system will help safeguard nature, but they want more transparency about the zones chosen for conservation.
- **Saving the iconic Colorado River — scientists say latest plan is not enough** [26 May 2023]
  • US states pledge to cut their use of water from the river, which has been drained by climate change and drought.
- **Finding the Amazon's tallest trees — an epic quest to reach hidden giants** [07 Jun 2023]

• When researchers spotted clues that trees were growing to record heights in the rainforest, they tried, failed and tried again to reach the remote site.

- NEWS
- 30 May 2023
- Correction 05 June 2023

Turkey's researchers fear loss of freedom after Erdoğan re-elected

Researchers expect the administration to further restrict autonomy and free speech. Some say they will move away or retire.

- Miryam Naddaf



Turkey has re-elected the incumbent Justice and Development Party government for another five years.Credit: Getty

Researchers have told *Nature* they fear that the re-election of Turkish President Recep Tayyip Erdoğan will lead to more restrictions on academic freedom. His administration has sacked thousands of university staff members, appointed unelected university rectors and curbed academic freedom and university autonomy since it came to power 21 years ago.

Erdoğan, leader of the governing Justice and Development Party (AKP), defeated Kemal Kılıçdaroğlu of the opposition Nation Alliance by a margin of 52.1% to 47.9% in election results confirmed on 28 May. Independent observers from the Organization for Security and Co-operation in Europe said the election offered "genuine political alternatives" but gave an "unjustified advantage to the incumbent".

"You won't hear much from Turkish universities in the next five years," says Yasemin Yalım, a founding member of the Turkish Bioethics Association who is based in Ankara. Many scientists and university teachers who had hoped for change in the May elections will now consider early retirement or emigration, she adds.

Before the election, the European University Association (EUA) in Brussels had expressed its willingness to support and advise the future government on increasing the involvement of the university community in institutional decision-making. The EUA told *Nature* that, if asked, its offer still stands. An EUA report published in March placed Turkey at the bottom of a scorecard for academic freedom in European countries and territories.

In 2016, Erdoğan's administration carried out one of the largest nationwide purges in Turkey's history, after an attempted military coup against it. More than 150,000 people lost their jobs, including over 8,500 university staff members.

Since 2018, rectors of public universities have been appointed or dismissed by the president without faculty elections. Turkish higher education is unique in Europe in that a university's leadership is not in the hands of the university itself, according to the EUA report.

Researchers worry that their voices will now be further silenced, especially on concerns such as safeguarding academic freedom and improving

universities' autonomy.

Self-censorship

Since coming to power in 2002, the AKP has tightened the government's control over universities and research councils, for example by micro-managing university spending and senior appointments. Out of fear for their career progression, researchers tend to avoid teaching and research into the rights of Kurdish people, people from sexual and gender minorities (LGBT+), and the history of the 1915 Armenian genocide by the Ottoman Empire.

"The one-man way of ruling has permeated into all universities, even the most progressive ones," says Canan Atılgan, a biophysicist at Sabancı University in Istanbul.

Yalım used to teach bioethics at Ankara University but says she took early retirement in 2020 because of declining academic freedom. She says she was warned by academic colleagues to be careful after screening a short film for students depicting mistreatment of LGBT+ individuals in Turkey's health system. "I am an ethicist; I need to have freedom of speech. If I don't have it, there is no meaning of teaching. That's why I retired," says Yalım. In a 2020 survey of 365 academics, around one-third reported censoring themselves in academic publications, teaching and professional gatherings.

Council of fear

Academics have repeatedly called for reforms to a government-run body called the Council of Higher Education (YÖK). Established in 1981, after a military coup in 1980, YÖK approves teaching courses, decides who should be appointed to faculty positions and sets student numbers. "Institutional autonomy is something that's supposedly protected under Article 130 of the constitution. So, it's completely violated," says Esra Mungan, a psychologist at Boğaziçi University in Istanbul.

In December 2022, a group of academics submitted a report to members of Turkey's Parliament calling for YÖK to be replaced with a more supportive

administrative body. The opposition alliance agreed to this recommendation, but the AKP made no such promise.

“It’s not out of the ordinary for a higher-education system to have a buffer body or a regulator acting between the government and the education system,” says Monika Steinel, the EUA’s deputy secretary-general. But such bodies are expected to operate with fewer direct interventions on institutional issues and decisions, she adds. “Academic freedom is the basis for a researcher or a university teacher to do their job,” says Steinel.

Political scientist Nebi Miş agrees that “systematic changes” are needed to Turkey’s higher-education laws. Before the election, the AKP proposed to establish boards of trustees for public universities. “This is a step that will strengthen academic autonomy,” says Miş, who is director of political studies at the SETA Foundation, a think tank based in Ankara with long-standing ties to the AKP government.

However, Mungan is not confident that these boards will protect academic autonomy, given the AKP government’s previous history. “There’s absolutely no chance that any oppositional figures like the ones who voted against Erdoğan would be serving on these councils.”

Nature contacted both Turkey’s science council, TÜBİTAK, and YÖK. The science council said it was unavailable for an interview. YÖK did not respond to *Nature*’s interview questions.

Stay and fight

Many researchers have left Turkey and others are expected to leave. But those to whom *Nature* spoke for this article say they are going to stay and fight for autonomy. They include Mungan, who spent 40 days in jail in 2016 for signing a petition condemning the Turkish government for attacking Kurdish provinces in southeastern Turkey.

Mungan has had offers to work abroad, but says she is not going anywhere. “I stayed in Turkey because we are millions, and [the election] showed that we are over 25 million that are not just against, but extremely against what is going on in our country.”

Nature **618**, 219-220 (2023)

doi: https://doi.org/10.1038/d41586-023-01775-7

Updates & Corrections

- **Correction 05 June 2023**: This article has been amended to make clear that Turkey’s military coup was in 1980, and not 1981.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01775-7

- NEWS
- 26 May 2023

US will vaccinate birds against avian flu for first time — what researchers think

The country plans to vaccinate endangered condors in an effort to curb unprecedented H5N1 outbreaks.

- Max Kozlov



US authorities have authorized the vaccination of California condors in a bid to protect them from a highly pathogenic strain of bird flu.Credit: Beth

Morris/Getty

US officials have authorized the vaccination of the critically endangered California condor (*Gymnogyps californianus*) against a type of avian flu spreading globally. It is the first time that the United States has approved inoculation of any bird against highly pathogenic avian influenza (HPAI).

The authorization comes as an H5N1 strain of avian influenza, a type of HPAI, has spread to an unprecedented number of countries, lasted longer than a typical outbreak and killed hundreds of millions of birds, as well as many mammals, worldwide. Some countries already vaccinate birds, including commercial flocks, against avian flu. The severity of the outbreak is driving some nations that have been hesitant, including the United States, to follow suit.



[Why unprecedented bird flu outbreaks sweeping the world are concerning scientists]

The decision to vaccinate California condors, a closely monitored species, does not mean that authorities are planning to vaccinate the country's poultry stock. Still, researchers will be watching how the inoculation campaign goes, with an eye towards administering the vaccines more widely.

"This authorization opens the opportunity to add another tool for how we address this threat," says Ashleigh Blackford, the California-condor

coordinator at the US Fish and Wildlife Service (FWS), headquartered in Washington DC.

Persistent flare-ups

The strain of HPAI that is currently spreading first infected birds in China in 1996 and has since caused sporadic outbreaks that usually flare up in the autumn and fizzle out by the following spring. But since late 2021, outbreaks have persisted through the summer in a record number of bird species, along with mammals such as raccoons, foxes and seals.

H5N1 has killed birds before, but “never to the extent we’re seeing in North America and Europe now”, says Samantha Gibbs, a wildlife veterinarian at the FWS, who specializes in avian influenza.

Since October 2021, more than 70 countries on five continents have reported outbreaks. And on 22 May, Brazil, which is the world’s largest poultry exporter, became the eighth South American country to declare an emergency, in response to the country’s first detection of HPAI in wild birds.

H5N1 can infect people, but cases are rare and typically involve close contact with infected poultry. No documented cases of human-to-human transmission of this strain of the virus have occurred thus far, but researchers worry that, with more opportunities to infect mammals, the virus could eventually evolve the ability to spread among people.

Condor conservation

Several vaccines against H5N1 are available, but administering the jab to birds has long been controversial among researchers and farmers. Poultry producers worry about the cost and difficulty of vaccinating millions of birds, as well as trade restrictions: many countries forbid the import of vaccinated birds owing to concerns that they could make it difficult to track the virus’s spread.



Avian flu freezes coastal bird research in South Africa

Until this outbreak, US authorities had claimed that strict biosafety protocols, such as enhanced disinfection procedures for farm workers or wildlife managers, and culling infected birds were enough to mitigate the worst effects of HPAI. The high mortality rate of the current strain of H5N1 in the California condor, which is the largest bird in North America, has forced health officials to re-evaluate this approach.

The species had a brush with extinction in 1987: only 27 wild condors remained at that time, so US officials captured them, bred them in captivity and released them back into the wild. Today, the birds remain one of the world's rarest avian species: as of last year, there were 537 condors, and 63% of them lived in the wild.

So far this year, officials have found 21 dead condors, 15 of which tested positive for HPAI. (The bodies of four birds couldn't be recovered, so their HPAI status is unknown.) It took more than 20 years to rebuild the population to its current level, Blackford says. "This is not the direction we want to be going in."



A poultry-vaccination campaign in France — which will include ducks that are farmed for foie gras — is due to begin in September.Credit: Stephane Mahe/Reuters

Inoculating these birds is a “perfect” use of the vaccine, says Karen Grogan, a clinical poultry veterinarian at the University of Georgia in Athens. But it’s unclear whether it will be effective in the condors, because it was licensed for use in chickens. “Even between chickens and turkeys,” which are a part of the same family of birds, “a vaccination might be efficacious in one but not the other”, she says.

Last week, scientists gave the jab to 20 black vultures (*Coragyps atratus*), which are part of the same family as the condors but are not endangered, to test the vaccine’s safety and efficacy. They will also vaccinate captive California condors before inoculating the wild population during an annual routine health check, Blackford says. The jabs use an inactivated form of an earlier H5N1 strain.

Stop the spread

This limited vaccination campaign in the United States is just one part of a broader movement among researchers to study the safety and efficacy of vaccination against avian influenza in birds.

The objective for the condor programme is to conserve an endangered species, says Jean-Luc Guérin, an avian pathologist at the National Veterinary School of Toulouse, France. In other regions, there are aims to curb the spread of the virus altogether.



Bird flu outbreak in mink sparks concern about spread in people

For example, the French government has ordered 80 million HPAI jabs as part of a mass-vaccination campaign for the country's poultry stocks, starting with ducks. This campaign, the first of its kind in the European Union, is due to start in September. China, meanwhile, has been vaccinating its poultry stock for nearly 20 years and has had some success in reducing infections in poultry and people[1].

Although the US Department of Agriculture is testing four vaccine candidates against H5N1 for use in poultry, it has not signalled that it plans to deploy a wider vaccination campaign for commercially farmed birds.

Gibbs says that, although vaccinations can be an important tool to keep outbreaks in check, the fundamentals of conservation work will be even more important to protect the health of wildlife, and ultimately of people.

Efforts to preserve and restore bird habitats will ensure that animals “can fight off not just bird flu, but whatever is going to come next”, Gibbs says. “If we set them up for success through conservation efforts, then we don’t have to play this catch-up game with pharmaceuticals.”

Nature **618**, 220-221 (2023)

doi: https://doi.org/10.1038/d41586-023-01760-0

References

1. Wu, J. *et al. Emerg. Infect. Dis.* **25**, 116–118 (2019).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01760-0

- NEWS
- 25 May 2023

Chronic stress can inflame the gut — now scientists know why

Signals originating in the brain make their way to gut nerve cells, leading to a release of inflammatory chemicals.

- Saima Sidik



Intestinal tissue (artificially coloured) of a person with ulcerative colitis.Credit: Steve Gschmeissner/Science Photo Library

Psychological stress is known to worsen the gut inflammation caused by certain bowel diseases. Now scientists have found out why. New research[1] outlines a sweeping narrative that begins with chemical cues produced in the brain and ends with immune cells in the gut — a sequence that spells trouble for people with these conditions.

The work, published today in *Cell*, helps to explain how chronic stress can trigger physical distress. And it implies that managing stress levels might have a profound influence over the effectiveness of treatments for inflammatory bowel disease (IBD). That idea runs contrary to conventional medical treatment, which has "completely neglected the psychological state of a patient as a major driver of [the] response to treatment", says study co-author Christoph Thaiss, a microbiologist at the University of Pennsylvania in Philadelphia.

The path from brain to gut

Abdominal pain, diarrhoea and fatigue are just a few of the symptoms that people with IBD experience. The two main types of IBD, ulcerative colitis and Crohn's disease, are mild in some people but, in others, can be debilitating or even life-threatening.

Stressful events, such as losing one's job or breaking up with a partner, often precede IBD flare-ups. Thaiss and his colleagues have now traced that linkage. After a surge of stress, the brain sends signals to the adrenal glands, which release chemicals called glucocorticoids to the rest of the body.



Anxiety can be created by the body, mouse heart study suggests

Initially, the researchers considered the idea that glucocorticoids act directly on immune cells, which respond by releasing molecules that cause inflammation. “But it turns out that there is a sort of layer in between,” Thaiss says. Working in mice, they found that glucocorticoids act instead on neurons in the gut and on cells called glia that connect gut neurons to one another.

Co-opted immune cells

After being switched on by glucocorticoids, some glial cells release molecules that trigger immune cells. In turn, those immune cells release molecules that would normally be used to fight off pathogens, but in this case end up causing painful bowel inflammation.

At the same time, glucocorticoids block immature gut neurons from developing fully, the researchers found. As a consequence, these neurons produce only low levels of signalling molecules that cause gut muscles to contract. This means food moves slowly through the digestive system, which adds to the discomfort of IBD.

The researchers were surprised to learn that glucocorticoids cause gut inflammation, because these compounds are sometimes used to treat IBD.

This apparent paradox might be explained by the short time frame on which such treatments are used. Although quick bursts of glucocorticoids seem to be anti-inflammatory, when stress becomes chronic, "the system completely shifts" and glucocorticoids take on a pro-inflammatory role, Thaiss says. It's a "plausible explanation", says gastroenterologist and immunologist John Chang at the University of California, San Diego.

Stress management for symptom relief

The brain's ability to drive inflammation in far-flung organs "seems to be much stronger" than was thought before, Thaiss says. This suggests that IBD drugs, in combination with stress-management techniques, could be more effective than the drugs alone. Molecules in the signalling pathway that runs from the brain to the gut could also become targets for new pharmacological treatments — "an exciting possibility", Chang says.

The implications of the work could reach beyond IBD. Stress is also thought to heighten inflammatory diseases of the skin and lungs, possibly through similar signalling pathways.

Moving forwards, Thaiss is excited to explore whether brain states other than stress influence a person's overall health. "There's definitely a huge amount we still need to learn about the brain and how the brain controls seemingly unrelated aspects of physiology and disease."

Nature **618**, 221-222 (2023)

doi: https://doi.org/10.1038/d41586-023-01700-y

References

1. Schneider, K. M. *et al. Cell* https://doi.org/10.1016/j.cell.2023.05.001 (2023).

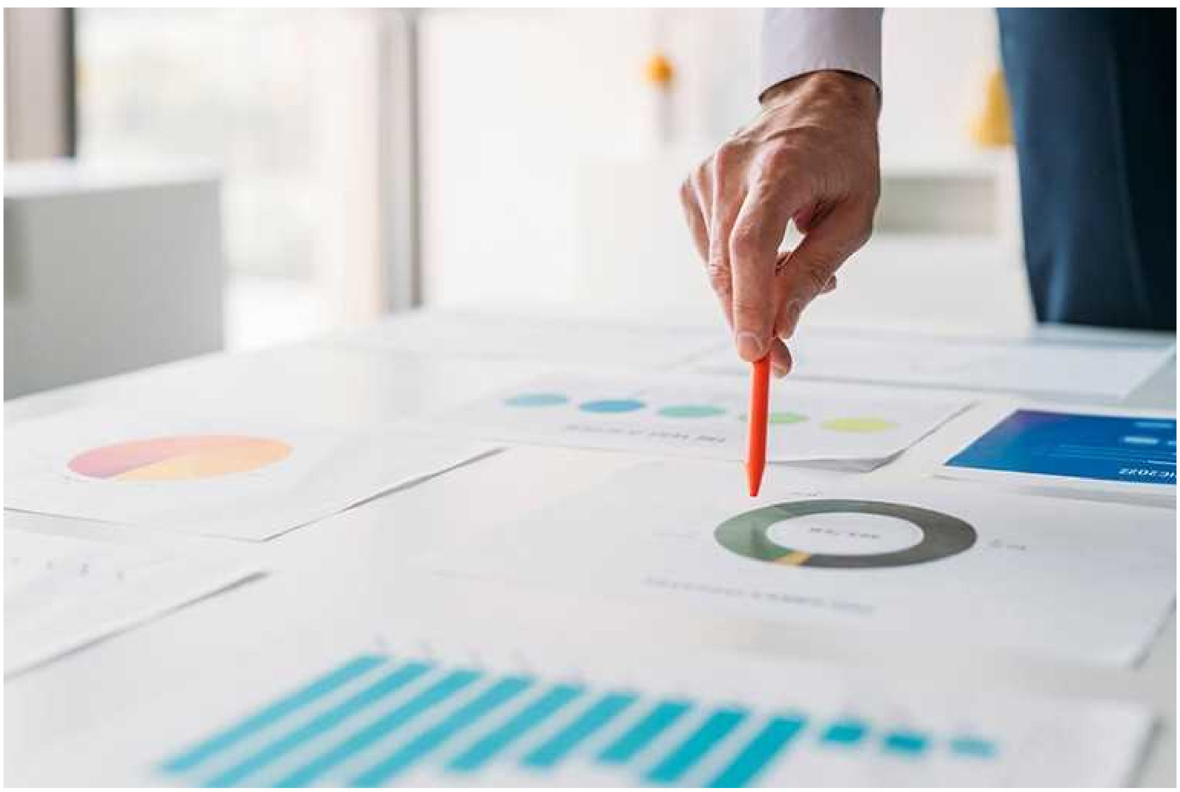
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01700-y

- NEWS
- 31 May 2023

AI intensifies fight against 'paper mills' that churn out fake research

Text- and image-generating tools present a new hurdle for efforts to tackle the growing number of fake papers making their way into the academic literature.

- Layal Liverpool



Artificial-intelligence tools that can generate realistic images and text are making it difficult to detect fake research.Credit: Westend61/Alamy

Advances in artificial intelligence (AI) are complicating publishers' efforts to tackle the growing problem of paper mills — companies that produce fake scientific papers to order. Generative AI tools, including chatbots such as ChatGPT and image-generating software, provide new ways of producing paper-mill content, which could prove particularly difficult to detect. These were among the challenges discussed by research-integrity experts at a summit on 24 May, which focused on the paper-mill problem.



Paper-mill detector put to the test in push to stamp out fake science

"The capacity of paper mills to generate increasingly plausible raw data is just going to be skyrocketing with AI," says Jennifer Byrne, a molecular biologist and publication-integrity researcher at New South Wales Health Pathology and the University of Sydney in Australia.

"I have seen fake microscopy images that were just generated by AI," says Jana Christopher, an image-data-integrity analyst at the publisher FEBS Press in Heidelberg, Germany. But being able to prove beyond suspicion that images are AI-generated remains a challenge, she says.

Language-generating AI tools such as ChatGPT pose a similar problem. "As soon as you have something that can show that something's generated by ChatGPT, there'll be some other tool to scramble that," says Christopher.

A stream of papers

Anna Abalkina, a social scientist at the Free University of Berlin and an independent research-integrity analyst, suspects that there might be a delay in these AI tools becoming more apparent in the academic literature because of the length of the peer-review process. Perhaps in the next few months, “we will see the first stream of papers”, she says.

Byrne, Christopher and Abalkina were participants at the UNITED2ACT summit last week, which was convened by the Committee on Publication Ethics (COPE), a non-profit organization focused on ethics in academic publishing based in Eastleigh, UK, and the International Association of Scientific, Technical and Medical Publishers (STM), based in Oxford. The summit brought together international researchers, including independent research-integrity analysts, as well as representatives from funding bodies and publishers.



The fight against fake-paper factories that churn out sham science

“It was the first time we had a group of people come together and co-create a set of actions which we’re going to take forward to combat this problem,” says Deborah Kahn, a trustee of COPE and a research-integrity consultant, based in London. The group intends to publish its joint action plan soon.

When it comes to detecting paper-mill works, “there is absolutely an additional challenge which is posed by synthetic images, synthetic text, et cetera”, says Joris van Rossum, programme director for STM Solutions, a

subsidiary of STM. “There is a general realization that there is the potential of screening becoming more difficult,” he says.

AI assistance

Kahn says that, although there will undoubtedly be positive uses of AI to support researchers writing papers, it will still be necessary to distinguish between legitimate papers written with AI and those that have been completely fabricated. “We have to really look at how we identify those things, and how we make sure that people have actually done the research. And there are various ways we can do that,” she says.

One strategy discussed during the summit was to require authors to provide the raw data from experiments, potentially with digital watermarks that would enable publishers to confirm that those data are genuine.



Hundreds of gibberish papers still lurk in the scientific literature

Currently, requirements for submitting raw data vary significantly between publishers, says Christopher. Establishing a uniform set of requirements for the submission of raw data across publishers, taking into account differences between fields of research, could therefore be helpful, she says.

Sabina Alam, director of publishing ethics and integrity at Taylor & Francis, a publisher based in Abingdon, UK, agrees but says that such standards will take time to implement. “I can’t imagine it being an overnight flip, because the reality is many institutions don’t actually have the resources to offer

data-management infrastructure," she says. "We don't want to penalize actual research."

Sharing information

The summit also discussed other strategies for tackling the problem of paper mills more broadly, including organizing an awareness day or week for researchers, as well as identifying ways for publishers to share relevant information on suspected paper mills — for example when publishers simultaneously receive submissions — without breaching data-protection rules.

STM is continuing to develop its own paper-mill detection software, while also collating resources on similar tools available elsewhere through its integrity hub. The apparent rise in paper mills increases demand for such techniques — both for detecting fake papers at the point of submission and for identifying those that are already published.

Taylor & Francis is among the publishers that are making use of such tools, and Alam says that a growing number of ethics cases — instances of potential misconduct that are flagged for further investigation — are being escalated to her team. Roughly half of these cases are because of paper mills, according to Alam. Her team saw the number of ethics cases increase more than tenfold from 2019 to 2022 — and so far this year, there have been almost as many cases as during the whole of 2022. "It seems to have been commercialized and scaled up," she says.

Nature **618**, 222-223 (2023)

doi: https://doi.org/10.1038/d41586-023-01780-w

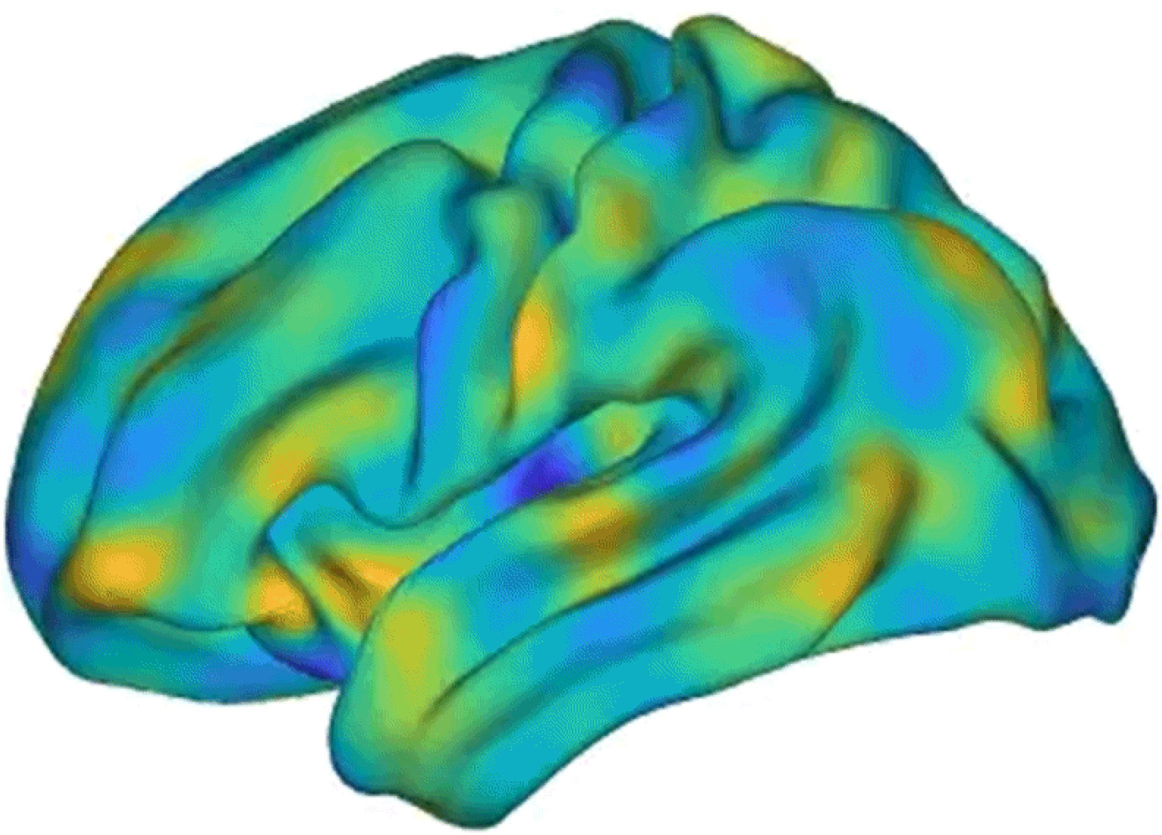
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01780-w

- NEWS
- 31 May 2023

The human brain's characteristic wrinkles help to drive how it works

A model of the brain's geometry better explains neuronal activity than a model based on the 'connectome'.

- Davide Castelvecchi



Waves of electrical activity spread across the resting brain in this simulation.J. C. Pang *et al./Nature*

The wrinkles that give the human brain its familiar walnut-like appearance have a large effect on brain activity, in much the same way that the shape of a bell determines the quality of its sound, a study suggests[1]. The findings run counter to a commonly held theory about which aspect of brain anatomy drives function.

The study's authors compared the influence of two components of the brain's physical structure: the outer folds of the cerebral cortex — the area where most higher-level brain activity occurs — and the connectome, the web of nerves that links distinct regions of the cerebral cortex. The team found that the shape of the outer surface was a better predictor of brainwave data than was the connectome, contrary to the paradigm that the connectome has the dominant role in driving brain activity. "We use concepts from physics and engineering to study how anatomy determines function," says study co-author James Pang, a physicist at Monash University in Melbourne, Australia. The results were published in *Nature* on 31 May[1].

Brain chatter

'Exciting' a neuron makes it fire, which sends a message zipping to other neurons. Excited neurons in the cerebral cortex can communicate their state of excitation to their immediate neighbours on the surface.

But each neuron also has a long filament called an axon that connects it to a faraway region within or beyond the cortex, allowing neurons to send excitatory messages to distant brain cells. In the past two decades, neuroscientists have painstakingly mapped this web of connections — the connectome — in a raft of organisms, including humans.



Famous ‘homunculus’ brain map redrawn to include complex movements

The authors wanted to understand how brain activity is affected by each of the ways in which neuronal excitation can spread: across the brain’s surface or through distant interconnections. To do so, the researchers — who have backgrounds in physics and neuroscience — tapped into the mathematical theory of waves.

Natural phenomena include many types of wave, from the seismic waves that constitute earthquakes to the electromagnetic waves of light. But all of these phenomena are captured by the same simple mathematical equation. This equation enables researchers to calculate the patterns of waves across a surface based exclusively on its geometry. These patterns can be broken down into fundamental components called modes, which are affected by an object’s geometry.

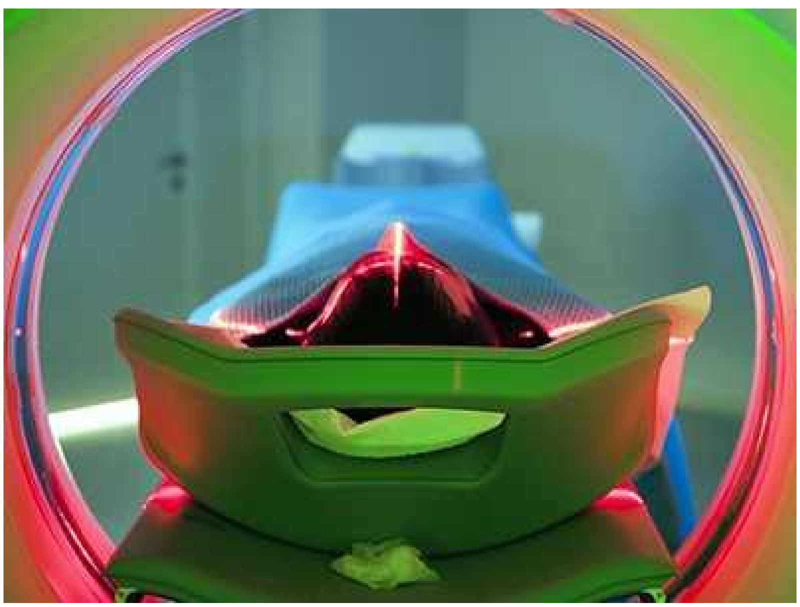
Good vibrations

For waves of vibrations propagating across a sphere, for example, the basic modes include a bulge moving back and forth from the top hemisphere to the bottom one and another travelling from left to right. But a bell lacks a lower hemisphere and therefore has modes different from those of a sphere. An object’s geometry also affects its modes’ characteristic frequencies and relative loudness.

The brain’s neuronal excitation can also come in waves, which can spread across the brain and travel back in periodic oscillations.

The researchers calculated the modes of brainwave propagation for the cortical surface and for the connectome. As a model of the connectome, they used information gathered from diffusion magnetic resonance imaging (MRI), which images brain anatomy. They then looked at data from more than 10,000 records of functional MRI, which images brain activity based on blood flow.

The analysis showed that brainwave modes in the resting brain as well as during a variety of activities — such as during the processing of visual stimuli — were better explained by the surface geometry model than by the connectome one, the researchers found.



[Your brain on psychedelics]

David Van Essen, a neuroscientist at Washington University in St. Louis, Missouri, who leads a connectome project, says that the diffusion MRI data the team used has well-documented drawbacks that make the comparison “not fair”. The team should have also looked at brain activity from simple stimuli that activate only local regions of the cortex, according to Van Essen. “It is extremely unlikely that the travelling wave model favoured by the authors could replicate such patterns,” he says.

Pang says that it would be interesting to test their models with such stimuli, and that the analyses he and his co-authors have done so far were a “proof of principle”.

For their study, the authors modelled an idealized brain structure, but the walnut-like convolutions of the cortex are known to vary in shape from one person to the next. The authors’ techniques could help to explore how such variations affect the corresponding modes.

Nature **618**, 223-224 (2023)

doi: https://doi.org/10.1038/d41586-023-01774-8

Read the related Research Briefing: The geometry of the human brain shapes its function.

References

1. Pang, J. C. *et al. Nature* https://doi.org/10.1038/s41586-023-06098-1 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01774-8

- NEWS
- 26 May 2023
- Correction 29 May 2023

China is using satellites to police the protection of nature — but will it work?

Scientists say the system will help safeguard nature, but they want more transparency about the zones chosen for conservation.

- Dyani Lewis



Golden snub-nosed monkeys (*Rhinopithecus roxellana*) are an endangered species, found only in China.Credit: Thomas Marent/Nature Picture Library

China's government has rolled out a nationwide satellite and automated monitoring system to protect land set aside for conservation from illegal development. Scientists hope that the move will safeguard ecologically important habitats and provide a model of remote-sensing use for conservation that other countries could follow. But they also have questions about how the nation has decided which areas to protect and where the boundaries, known as the ecological redlines, lie.

"The decision makers have made a really bold step forward," says Chi-Yeung Choi, an applied ecologist at Duke Kunshan University in Suzhou, China. He says that having a national system to protect ecologically important areas will stop provincial governments prioritizing development over conservation. The policy "has really huge potential to conserve biodiversity hotspots", he says.

As a vast nation, China spans diverse habitats such as grasslands, forests, deserts and mountains. It is home to iconic species including the giant panda (*Ailuropoda melanoleuca*), south China tiger (*Panthera tigris tigris*) and snub-nose monkeys (*Rhinopithecus* spp.), as well as tens of thousands of lesser-known species. The International Union for Conservation of Nature (IUCN) counts it as the third most biodiverse country in the world.

On 22 April, China's Ministry of Natural Resources announced that its ecological-redline map was finalized. Conservation areas in each mainland province are now organized into a central system managed by the ministry. Previously, China's conservation zones included 2,750 nature reserves and thousands of other areas protected by different levels of government.

According to the announcement, the protected zones cover 3 million square kilometres of land — approximately 30% of the Chinese mainland — and 150,000 square kilometres of sea. That's in line with the 2022 Kunming–Montreal Global Biodiversity Framework target to have at least 30% of terrestrial and inland water areas conserved by 2030. Zhijun Ma, a conservation biologist at Fudan University in Shanghai, China, says that the redline map doubles the area that is legally protected from development.

Lack of transparency

Four days after the redline map was finished, the Ministry for Ecology and Environment announced the launch of a monitoring platform to police the network of zones. It said that a fleet of 30 satellites has been launched to capture high-resolution images, with an algorithm designed to automatically detect changes to forest cover and land use within the redlines.

The satellites are the space component of a three-layered monitoring network, referred to as ‘space, sky and land’, says Fangyuan Hua, a conservation biologist at Peking University in Beijing. Sky and land refer to drones and on-ground personnel who investigate human activity in the protected zones. The aim is to enable authorities to act swiftly if they detect illegal activity, such as land clearance for new mines or property development.

But environmental scientists keen to research how effective the zoning will be are frustrated by a lack of transparency from the government. “Even though the map has been finished across the country, we actually don’t know what the map looks like,” says Hua. Without a public record of the conservation boundaries “there’s a possibility that the local governments might be able to shift their redlines to accommodate future [development] needs”, making it impossible for the public to monitor what’s going on, she says.

Last year, Choi and his colleagues used a draft of the map to show that three times as many coastal sites that were important for waterbird conservation — around 75% of 172 sites — would be protected by the new plan than under its predecessor, the national nature reserve system[1]. “Assuming that things will go as planned, then there’s actually a huge increase in the number of sites and the amount of areas that can be protected,” he says. But Hua says that without access to final versions of the map, few studies such as Choi’s will be able to assess redline protections.

Alice Hughes, a conservation biologist at the University of Hong Kong, says that other nations could follow China’s example, but “China also should take steps to make the data it is collecting more available”.

China’s Ministry of Natural Resources did not respond to *Nature*’s request for comment on when the maps will be publicly available.

Ecosystem services

China’s approach to defining its redlines differs from how many countries set aside protected land. Generally, decisions are based on an area’s value to wildlife or plants. The IUCN divides protected areas into seven categories that differ in size and level of protection. China’s redline map includes habitats that are identified as being important for protecting vulnerable species and ecosystems, but also draws on assessments of ecosystem services, a measure of an area’s benefits to humans. For example, the metric attributes value to areas with ecosystems that sequester carbon (such as trees), store or purify water, prevent soil erosion or desertification, and protect biodiversity. Including ecosystem services as a criterion for protection “is a very uniquely Chinese thing”, says Hua.

Using these measures to justify land protection might be appealing to local governments, says Hughes, because preventing events such as landslides or sedimentation in waterways brings financial benefits. Overlapping conservation and economic priorities can lead to “mutual benefits”, she says.

But areas worth protecting for their human benefit might not overlap with areas worth protecting for their biodiversity, says Hua. And whereas ecosystem services can be readily assessed using remote-sensing data — to evaluate canopy cover for carbon sequestration, for example — the same isn’t true for mapping biodiversity, she says. That requires greater on-the-ground assessments, which are more expensive to conduct.

Hughes says that more ecological surveys are needed anyway, because detailed data on species vulnerability and distribution are lacking in China. As a result, conservation areas tend to favour large, charismatic animals such as the panda and snow leopard (*Panthera uncia*), and overlook less-celebrated species, especially in the southern parts of the country, for which the data are limited.

Unlike most of the IUCN categories of protected areas, which strictly limit human activity, the redline zones often allow some human activity within the borders, says Hua. For example, people might be permitted to live or cultivate crops in certain areas. “The network of redlines is mostly intended to prevent large-scale developments, like mining and construction,” she says.

The satellite monitoring system will be “a very powerful and effective system, if it can be done properly”, Choi says. And if the information on it is made open, researchers will also be able to interrogate land-use changes across the country.

Nature **618**, 224-225 (2023)

doi: https://doi.org/10.1038/d41586-023-01714-6

Updates & Corrections

- **Correction 29 May 2023**: This article originally stated that China is the first nation to use satellite monitoring. China is the first to use it in combination with machine learning. The text has been corrected.

References

1. Choi, C.-Y. *et al. Conserv. Lett.* **15**, e12853 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01714-6

- NEWS
- 26 May 2023

Saving the iconic Colorado River — scientists say latest plan is not enough

US states pledge to cut their use of water from the river, which has been drained by climate change and drought.

- Alexandra Witze



Visitors to Lake Mead, a reservoir in Nevada and Arizona that is fed by the Colorado River, see a bleached bathtub-like ring on its banks, indicating past

water levels.Credit: Justin Sullivan/Getty

Seven US states this week finally agreed on a plan to cut their use of water from the Colorado River, an icon in the nation's Southwest that supplies water to more than 40 million people, including in Mexico. The river, battered by overuse, drought and climate change, has been drying up. Although scientists welcome the plan, they say the agreement is only a temporary fix to a much thornier problem.

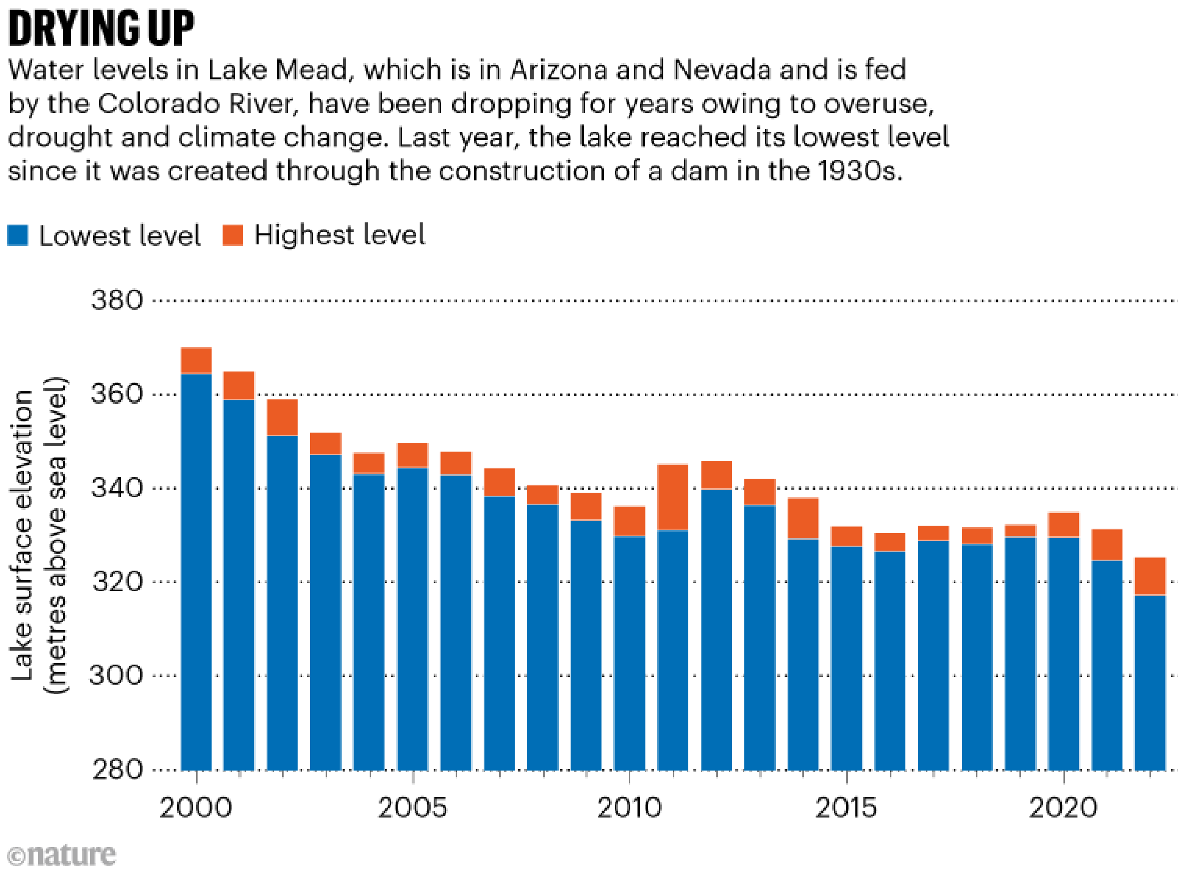


The world faces a water crisis — 4 powerful charts show how

"It does not change the fundamental problem of the overallocation of the Colorado River," says Kathryn Sorensen, an economist at Arizona State University in Phoenix. "But it helps." Combined with an unexpectedly wet winter in the western United States, which is now filling the river with snowmelt, the agreement buys time for officials to negotiate a more sustainable solution.

From its headwaters in the Rocky Mountains to where it empties into the Gulf of California, the 2,300-kilometre-long river serves farmers, homeowners, businesses and many others, including people in 30 tribal nations and in cities including Los Angeles. The 22 May announcement is the latest in a decades-long dispute over who should cut back on their water use to preserve the river. The 'lower basin' states of California, Arizona and Nevada — the three that are farthest downstream as Colorado flows to the sea, with California the heaviest user of the river's water — have agreed to

conserve 3 million acre-feet (3.7 trillion litres) of water between now and 2026. That amount is enough to supply around 6 million households for a year.



Source: US Bureau of Reclamation

Researchers say that the Colorado River is overtaxed by serving more people than it can handle, in a part of the country that doesn't get a lot of rainfall. A drought that began in 2000 has caused the river's water levels to drop steadily — culminating last year when the massive Lake Mead reservoir in Nevada and Arizona shrank (see 'Drying up') to the point that long-lost human bodies emerged on the shoreline. If the levels in Lake Mead and in Lake Powell in Arizona and Utah — both fed by the river — drop low enough, water will no longer pass through their dams to generate the thousands of megawatts of hydropower that are used by people across the west. Both reservoirs are currently only about 30% full, down from 95% full in 2000.

The inevitability of climate change

One bright spot in the Colorado River outlook is this winter's heavy snows, which are now melting from the mountains and feeding the river. As a result, lakes Powell and Mead are starting to rise faster than they have in recent years. But scientists say that it is a rare event, and not something to be counted on in the future. "This one year doesn't get us out of it," says Jack Schmidt, a geomorphologist at Utah State University in Logan.



What the science says about California's record–setting snow

Climate change, however, is plodding and relentless — and it's expected to increase the frequency of droughts in the Colorado River basin. Every temperature increase of 1 ºC in the upper part of the river basin leads to a 9.3% drop in the river's flow, the US Geological Survey has estimated[1]. Long-term strategies are needed to account for the inevitability of climate change, as well as the uncertainty of wet and dry seasons, says Kevin Wheeler, a water resource engineer at the University of Oxford, UK.

Reducing water use is a linchpin to all such plans. In California and Arizona, for instance, water management officials are ramping up plans to recycle wastewater into drinking water. Some are experimenting with encouraging changes in residents' behaviour by, for instance, charging more for water in

the summer than in the winter. Other efforts focus on cutting back on agricultural uses of water — which is key, because farmers account for 75–80% of the water consumption from the Colorado River, much of which is used to grow hay and other fodder for livestock. Technologies such as drip irrigation, which delivers water directly to plants rather than flooding an entire field, can help — although their effectiveness is debated.

A temporary fix

The four states that are upstream along the river — New Mexico, Utah, Colorado and Wyoming — have agreed with the lower-basin states. Together, they are asking the federal government, which oversees water releases from lakes Powell and Mead, to consider the new plan as a way forward between now and 2026. The US Department of the Interior has the power to modify, reject or accept the plan. The proposed cutback is only around half of what the department had been asking of the group of seven, but it still agreed to pay US$1.2 billion to support the proposal, for instance by paying farmers to let their fields lay fallow.



Mexico is seeding clouds to make rain — scientists aren't sure it works

That's a lot of money to support what is only a temporary fix, says Jennifer Gimbel, a water policy expert at Colorado State University in Fort Collins.

“It concerns me that we’re spending that kind of money when we should be focused on a more permanent solution,” she says. Any permanent solution will involve a complex stew of government-backed incentives to conserve water across the entire river basin, all trying to balance the needs of many users.

The agreement pushes off any further decision-making until 2026, when interim water-management guidelines are due to expire. Between now and then, the United States, Mexico and tribal nations must come up with an approach for how they will conserve the river’s limited resources in future.

There’s no time to waste, Sorensen says. “Everyone understands that time is of the essence.”

Nature **618**, 225-226 (2023)

doi: https://doi.org/10.1038/d41586-023-01746-y

References

1. Milly, P. C. D. & Dunne, K. A. *Science* **367**, 1252–1255 (2020).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01746-y

Ancient giants

An epic quest to reach the tallest known tree in the Amazon

By Richard Monastersky
Photography by Pablo Albarenga for *Nature*
7 June 2023

Wh
tri

Pab

Thi



“It’
tall
whi
spe

Gor
nor
eve
fuel
wer

“I w
Gor
Bra

The team packs up its boats.

The team packs up its boats.

Then it travels up the Jari River.

Then it travels up the Jari River.

In September last year, Gorgens and a team of 19 others returned, armed with key lessons from their first attempt, better equipment and supplies for a longer mission. A *Nature* photographer accompanied them on their journey.

Their expedition sounds at times like a tale from science in the nineteenth century — hacking through unexplored jungle with machetes, dodging deadly animals, enduring torrential rains. But they also had help from twenty-first-century tools and data. A drone gave views from high above the forest canopy, and the team had exact GPS coordinates for the unknown tree.

The researchers are exploring basic questions about what controls tree height. And the data they collected could help to fill in details about how much carbon the Amazon holds, and whether threats such as climate change and deforestation will disrupt that capacity. But beyond those questions, the expedition serves as a reminder of the vast realms of unexplored wilderness and science, even in a world made small by mobile phones, the Internet and other modern technology.

“It’s incredible when you see how small we are at that fantastic site,” says Gorgens. “Not just the tree. All. The river, the challenges, the size of the forest.”

A second chance

A drone shot shows some of the rapids the team had to navigate on its way up the Jari River.

A drone shot shows some of the rapids the team had to navigate on its way up the Jari River.

During the 2019 expedition, Gorgens and his crew managed to visit a cluster of big trees, but their heart-wrenching failure to reach the tallest one left them eager to return. Then the COVID-19 pandemic hit, and it wasn’t until 2022 that the researchers could mount another attempt.

This time, they had a much better sense of what they would run into. The first leg would be a 260-km trip up the Jari River from the small community of Laranjal do Jari, near the mouth of the Amazon River (see ‘Amazon adventure’). That would take four days. Then the hard part would begin: a 20-km slog through the same steep terrain and dense jungle that had forced them to turn back in 2019.

Starting in the community of Laranjal do Jari, a research team travelled more than 250 kilometres up the Jari River by boat and then trekked overland to reach the tallest known tree in the Amazon.

They began last year’s journey with 20 people, including 19 Brazilian scientists and crew members, and one UK researcher. But the problems started almost immediately. On the first night, one of the young team members from the local community came down with a fever and his leg started to swell. It seemed like symptoms of a spider bite, but he insisted he would be able to carry on. The same night, another crew member found out that his mother had died, so the next morning a boat ferried him several hours back to town and then returned. It cast a shadow over the expedition.

Several times over the next few days, the river was so low the crew had to get out of the boats and pull them through rapids. For one section, everyone unloaded the boats and carried them.

The team had to pull the boat through some rapids. (This video has no sound)

The team had to pull the boat through some rapids. (This video has no sound)

Meanwhile, the man with the spider bite developed a high fever and signs of an infection. “He started to decline in the first two days quite dramatically,” says Gorgens. The crew nearly turned around to evacuate him before he responded to a new course of medication. Finally, after four days on the river, the team could start the second stage of its quest.

The researchers and support crew set up a base camp by the river, where three people remained, including the man with the spider bite. Gorgens and his team pored over maps and spent a day planning routes and organizing supplies. They set off for the tree around 6 a.m., hoping to get the most out of the daylight.

Captured by laser

At this point, Gorgens was in the odd position of knowing precise details about the tree – its height and position – without having seen it. It’s possible that no one had.

(No Indigenous people live within 100km of the tree, and even if they had passed by, it would have been impossible to see the top through the thick canopy.)

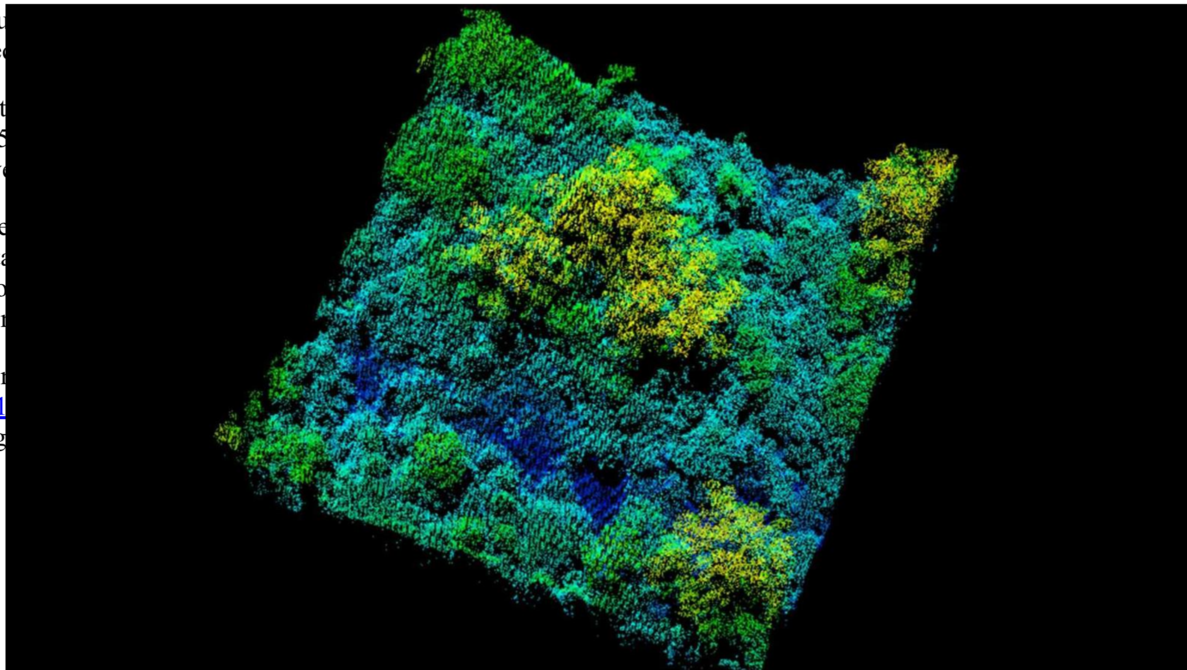
It was a chance discovery that brought Gorgens to this place. In 2018, he was studying height measurements of the forest collected by a Cessna aeroplane that had flown tracks over different parts of the Amazon between 2016 and 2018. The plane carried a lidar device that sends out laser pulses and measures the time it takes the reflections to

retu
prec

But
or 5
inv

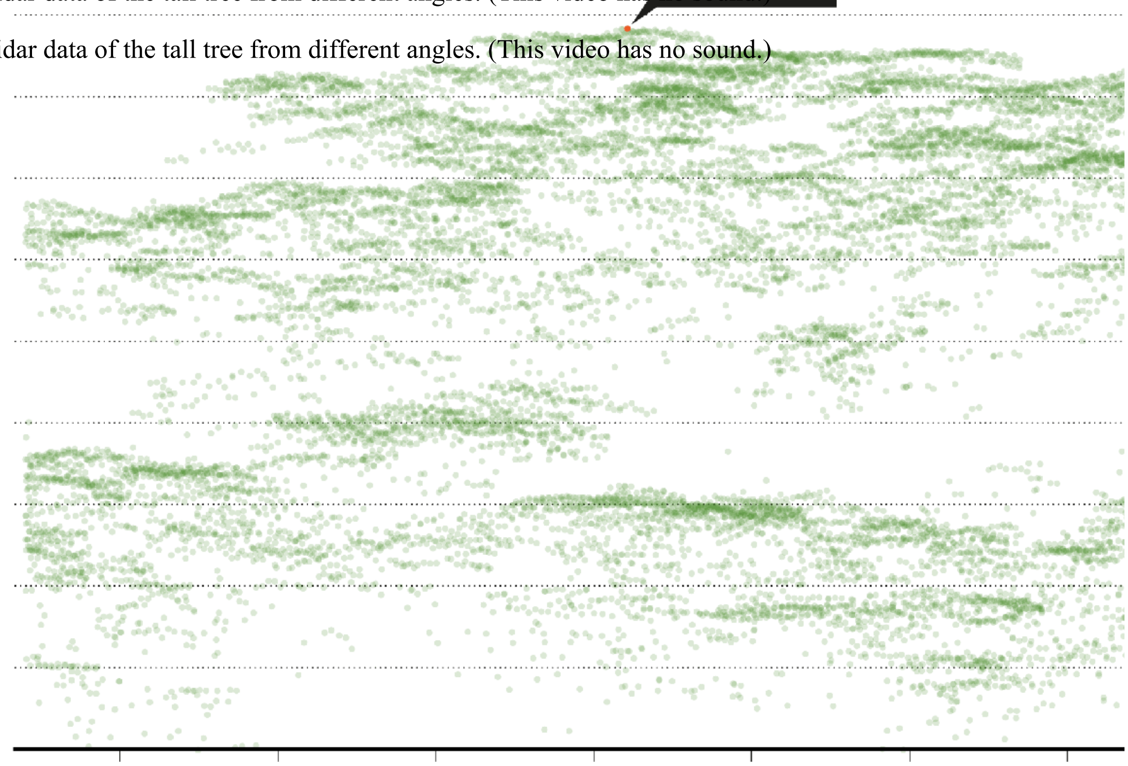
The
data
kno
wor

Gor
201
slig



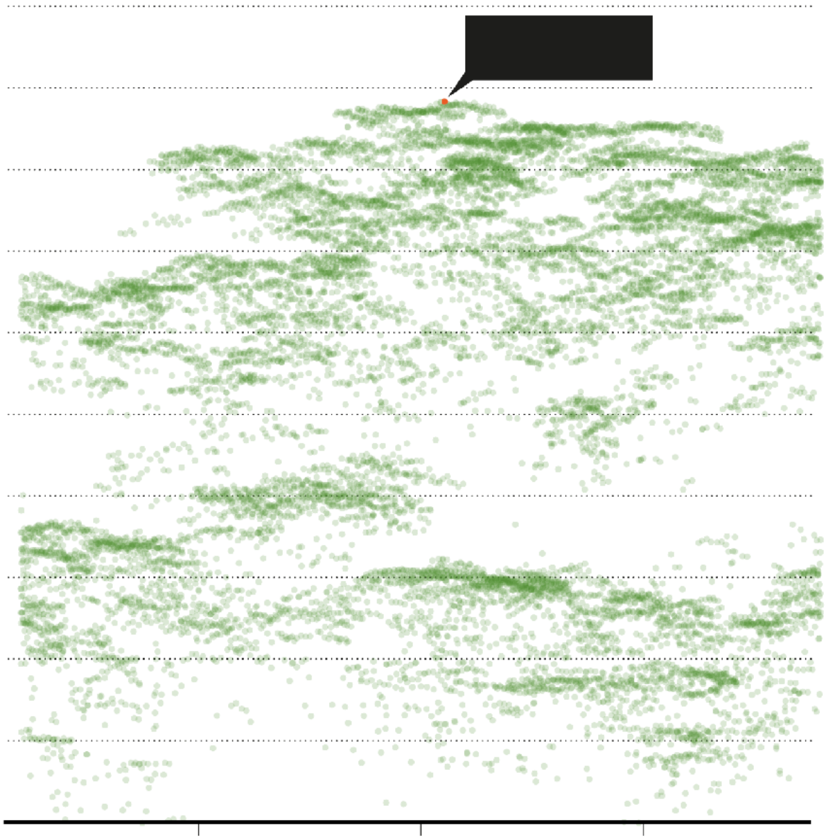
Airborne lidar data of the tall tree from different angles. (This video has no sound.)

Airborne lidar data of the tall tree from different angles. (This video has no sound.)



High life

Da
An ent
ref
10(
Tal
88.
90
80
70
Tree height (metres)
60
50
40
30
20
10
0
335
340
345
350
355
360
365



Mer(

So
Go

High life

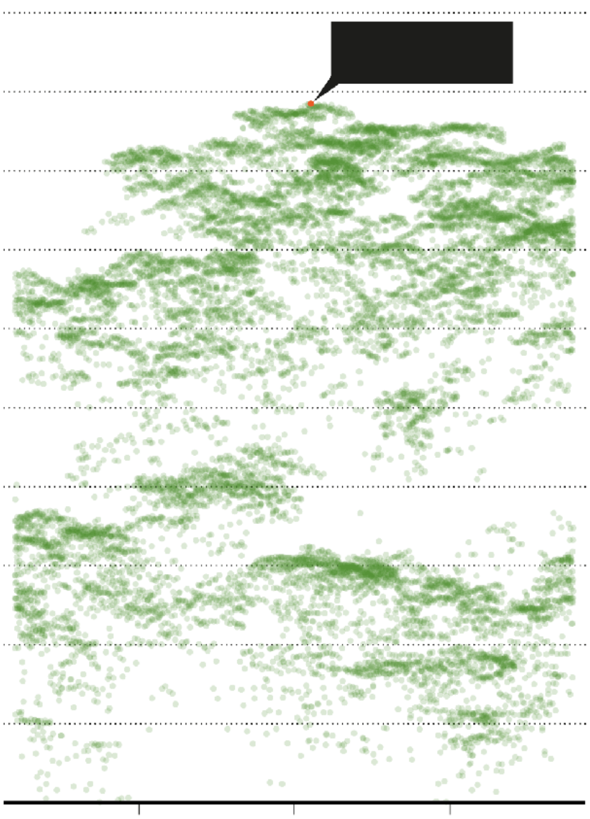
Data from an airborne lidar survey captured evidence of several extremely tall trees in the Amazon, which prompted the expedition to find and investigate one of them. Dots represent reflections of laser light off foliage.

100

Tallest tree

88.5 metres tall

90
80
70
e height (metres)
60



High life

Data from an airborne lidar survey captured evidence of several extremely tall trees in the Amazon, which prompted the expedition to find and investigate one of them. Dots represent reflections of laser light off foliage.

100

Tallest tree

88.5 metres tall

90

80

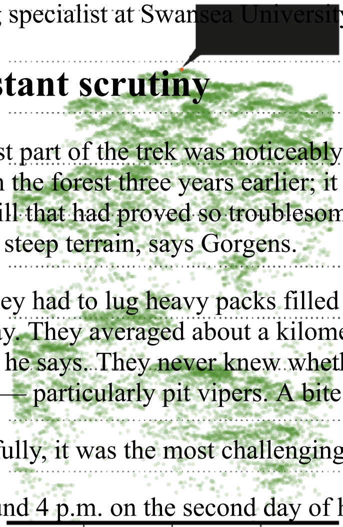
70

60

Height (metres)

When the researchers published their findings, including a lidar-produced profile image of the tree, they captured attention across the globe — particularly in Brazil. One political party even used the ghostly image of the tree in some of its work. "It's a symbol," says Gorgens. "To connect people to the environment."

For scientists, the lidar data helped to underscore how much remains unknown about the Amazon. "This discovery, from just a couple of years ago, highlights that we still don't know everything," says Jacqueline Rosette, a remote-sensing specialist at Swansea University, UK, who was part of last year's expedition.



Constant scrutiny

The first part of the trek was noticeably easier because Gorgens and his team were following the route they had cut through the forest three years earlier; it had only partially grown over. Then they took a wide route around the steep hill that had proved so troublesome during their previous attempt. With better maps, they were able to avoid a lot of steep terrain, says Gorgens.

Still, they had to lug heavy packs filled with food, tents and equipment in hot, humid conditions for 10–11 hours each day. They averaged about a kilometre per hour. "This is quite stressful because we don't know what will come," he says. They never knew whether they might hit some new obstacle, and were on constant watch for snakes — particularly pit vipers. A bite from one of the many venomous species in the area could prove deadly.

"Truthfully, it was the most challenging and physically arduous thing I've ever done," says Rosette.

At around 4 p.m. on the second day of hiking, they reached their destination. And there it was, a massive tree trunk about 10 metres around, at exactly the right GPS coordinates.

Still, initially Gorgens wasn't sure. "Thirty, forty metres away from this point, there were other big trees similar to this one," he says.

To confirm the tree's height, they would need a view from above the canopy. Although the team had brought climbing equipment during its 2019 expedition, they had decided it wasn't worth the extra weight this time. Instead, they had a drone to do the work.

Bird's-eye view

Once above the canopy, it was clear that the tree rises high above its neighbours. It belongs to the species *Dinizia excelsa*, also known as angelim vermelho, which are some of the tallest trees in South America. The other 80-metre-tall trees that Gorgens and his team have seen in the region are the same species.

From the vantage point of the drone, Gorgens could also see some of the surrounding topography. Close to the tree is a big hill. Rising some 200 m higher than the tree top, it might be one of the reasons the giants thrive here, perhaps providing some protection from the wind. "I think this hill, close to the tree — I think it helps to promote the tree to grow," says Gorgens.

Drone flight up the tallest tree. (This video has no sound.)

Drone flight up the tallest tree. (This video has no sound.)

That's just one possibility. Beyond that, the region doesn't get too much rain and has a lot of clouds, says Gorgens. "So we're proposing that especially with [less] wind and indirect light, you create a very calm environment allowing trees to grow for long periods without disturbance." They suspect the tree is at least 400 years old, but did not have permission to take a core sample to age it precisely.

The team collected leaves to deposit at a local herbarium, marked off a plot around the tree to catalogue the surrounding vegetation and collected soil samples from around the tree. Gorgens wants to do genetic analyses on these large *Dinizia* specimens to see whether they carry gene variants that differ from other *Dinizia* trees around the Amazon.

Rosette also used a handheld lidar device to scan the forest to obtain a high-resolution 3D structure of the vegetation surrounding the tree.

20
10
0

In the neighbourhood

,340
,350
,360

One surprise, she says, was the lack of smaller *Denizia* trees nearby that could replace the giant one after it dies. Another discovery came only later, after the team returned home. When the researchers calculated how much carbon was stored in and near the tree, they estimated that the giant trees hold about 60–70% of all the carbon in the surrounding hectare.

verse Mercator

The big trees "are very much the big boss of that surrounding area", says Rosette. "It was quite worrying actually to see you didn't have the youngsters kind of coming through, ready to exploit when these older, mature trees succumb to natural mortality."

Source: E. B. Gorgens

Nevertheless, the fact that trees can grow so tall in this part of the tropics changes how researchers understand the region and its ability to store carbon that would otherwise contribute to global warming, she says. "This is a vital and a new discovery."

The team camped at the tree's base to continue measurements the next morning, but got little sleep because an intense storm swept through the area. Some people spent hours holding up their tents, and stinging ants invaded the camp during the night. At one point, a viper slithered through and had to be killed.

The next morning, they finished their measurements and packed for the two-day trek back to the boats. It was much easier going because they followed the same path back. And the boat trip downriver took one less day because they travelled with the current.

Use the navigation arrows to explore a 360-degree view of the tree and the rainforest canopy.

A 360-degree view of the tree and the rainforest canopy. Credit: Pablo Albarenga

A 360-degree view of the tree and the rainforest canopy. Credit: Pablo Albarenga

Awaiting discovery

Gorgens knows he hasn't seen the tallest tree in the Amazon. "I'm sure we will find bigger trees," he says. The lidar surveys they did several years ago covered just 0.01% of the Amazon. So much of the basin remains unexplored, and he is hopeful that they can do more aerial surveys.

And the researchers now have access to another data set that offers clues for where to look for more tall trees. Between March 2019 and March this year, a lidar instrument called the Global Ecosystem Dynamics Investigation (GEDI) mission collected data from its position on the International Space Station. It has produced a map for the entire Amazon, but at much lower resolution than the data collected during plane-based surveys.

"We definitely see trees that are over 80 m in the Amazon," says Ralph Dubayah, principal investigator for GEDI and a remote-sensing scientist who studies ecology and climate at the University of Maryland in College Park. The lidar data collected from space even show some reflections over 90–95 m in the Amazon, but Dubayah says these could be artefacts. Aerial lidar would be the gold standard, he says: "You can get highly accurate estimates of height — better than you can get in any other way."

Dubayah adds that the remote monitoring has provided important information. "It's been revolutionary in giving us insights into where these important stands of individual trees are and what we can do to protect them."

The issue of protection for the Amazon is something the expedition members thought about as they made their long journey to and from the giant *Denizia*. The tree is in a protected zone, as are many of the other large trees.

"But it's not completely remote," says Gorgens. There is illegal mining close to some of the tall trees, he says. "We need to pay attention to these aggressions."

During the tenure of Brazilian president Jair Bolsonaro from 2019 to 2022, deforestation rates spiked in the Amazon as the government supported agricultural expansion, while [illegal mining and drug operations spread](). Forest advocates hope that the administration of current President Luiz Inácio Lula da Silva will bring protections back and rein in the disturbances.

Last week, state-government representatives started discussing how to protect these tall trees, says Diego Armando Silva da Silva, a forest engineer at the Federal Institute of Amapá in Laranjal do Jari, who was the local leader for the expedition.

Da Silva says that when he stood before these giant Amazon trees, "I felt like I was in front of a monument of nature." That's a common reaction from the scientists on the team.

Beyond their height, Gorgens thinks about how long the trees have stood in the forest. "They were here when the Portuguese arrived. They were here when the Indigenous people had to leave their lands. They were here when climate change started to increase the disturbances," he says. "Our challenge is to try to understand what they say about us. To understand ourselves through the eyes of the giant trees."

Ma
the



Na

doi

SP
Spr

- Terms & Conditions
- Accessibility statement

© 2023 Springer Nature Limited. All rights reserved.

TopBuilt with **Shorthand**

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01828-x

Books & Arts

- **Why we can thank a polio emergency for the birth of intensive care** [06 Jun 2023]
  • An outbreak of polio in 1950s Denmark led one hospital to pioneer mechanical ventilation, constant monitoring of vital signs and other innovations that are saving lives to this day.

- BOOK REVIEW
- 06 June 2023
- Correction 08 June 2023

Why we can thank a polio emergency for the birth of intensive care

An outbreak of polio in 1950s Denmark led one hospital to pioneer mechanical ventilation, constant monitoring of vital signs and other innovations that are saving lives to this day.

- Eric Topol



Bulky ‘iron lung’ devices were used to assist patients’ breathing in polio epidemics in the 1950s.Credit: Science History Images/Alamy

The Autumn Ghost: How the Battle Against a Polio Epidemic Revolutionized Modern Medical Care *Hannah Wunsch* Greystone Books (2023)

The COVID-19 pandemic has brought home the central role of intensive care units (ICUs) in saving the lives of those in critical condition in hospitals today. Yet if you asked most people where the ICU concept came from, few would know that it was an outgrowth of a polio epidemic in Denmark. In her brilliant new book, Hannah Wunsch, an anaesthesiologist and critical-care-medicine specialist at the University of Toronto, Canada, traces the origins of the modern ICU to 1952 and the Blegdam hospital in Copenhagen — something she has written about before in *Nature* (go.nature.com/45B6snd). There, a series of innovations arose out of dire need, including positive-pressure ventilation (the precursor to mechanical ventilators), blood-gas

measurements for pH and carbon-dioxide levels and close monitoring by an interdisciplinary team of nurses, doctors (notably anaesthesiologists), pharmacists and others.

The treatment of one patient, a 12-year-old girl named Vivi Ebert who presented with bulbar paralytic poliomyelitis — in which poliovirus infects the brainstem — forms the centrepiece of Wunsch's book. Of the first 31 people to be admitted to the Blegdam in the summer of 1952 with paralytic or respiratory polio symptoms, 87% died, 70% within three days. Thanks to interventions including manual ventilation, supervised by anaesthesiologist Bjørn Ibsen, Ebert survived another twenty years, eventually succumbing to pneumonia at the age of 32.



The outbreak that invented intensive care

In those early days it took 50 people to provide the muscle power required for round-the-clock ventilation for 6–8 people with paralytic polio. The hospital's initial success led to more than 1,500 medical and dental students being employed as manual ventilators for patients admitted in the summer and autumn of 1952. Eventually, mechanical ventilator machines took the place of humans, and the ICU concept was built, focusing on the sickest patients, who required a breathing machine and constant monitoring. Over the next few years, the use of ICUs expanded to the treatment of people with

major trauma, shock, tetanus and a variety of other acute, life-threatening conditions.

The treatment of polio, the main story of *The Autumn Ghost*, has rich parallels to the COVID-19 pandemic. In the 1950s, the prevailing hypothesis about the spread of polio was that the virus was inhaled into the body's upper airways. It took decades for the gastrointestinal transmission route — oral contact with the faeces of an infected person — to become accepted. Similarly, for COVID-19, there was an initial fixation on liquid droplets on surfaces and in the air as the main means of transmission, whereas it was determined later that it was spread predominantly within tiny droplets or aerosols in the air.

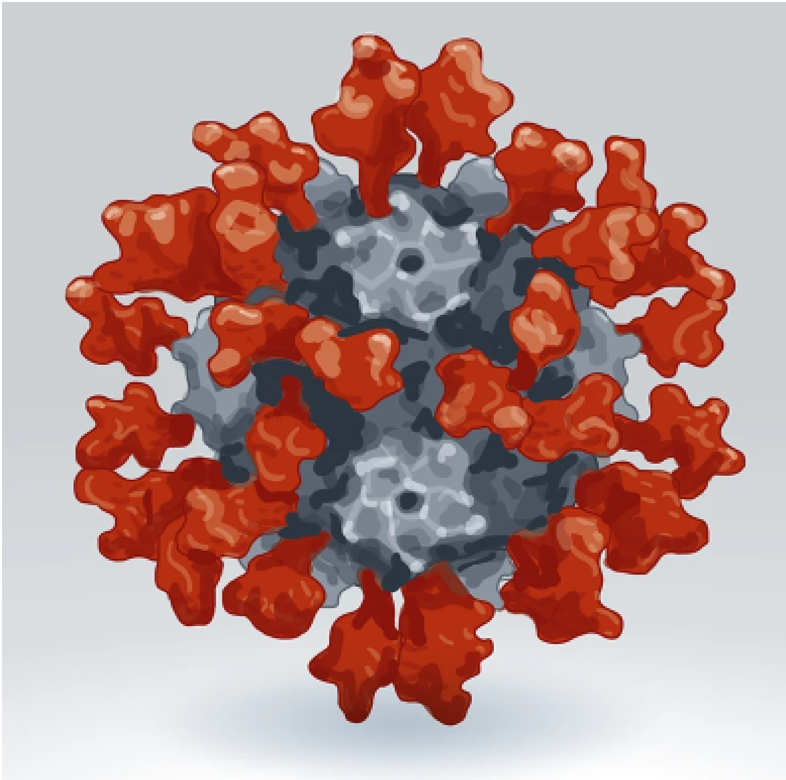
Furthermore, a substantial proportion of both poliovirus and SARS-CoV-2 infections were asymptomatic. And both viruses have long-term consequences: for polio, not only potential paralysis, but also the debilitating neuromuscular syndrome that can occur decades later. Long COVID affects 10–12% of infected individuals, with a variety of enduring symptoms that can be incapacitating with potentially more longer-term effects that are yet unknown.



Mechanical ventilators have been at the fore in the COVID-19 pandemic.Credit: Yasuyoshi Chiba/AFP via Getty

Polio taught us about the efficacy of positive-pressure ventilation for those having difficulty with breathing. With COVID-19, we learnt that ventilating patients while they were lying face down was crucial to good outcomes. For poliovirus, large, randomized trials of γ-globulin — a substance derived from bone marrow and lymph gland cells containing antibodies thought to help fight the virus — had some success in the years before a vaccine became available. For COVID-19, large observational studies were undertaken of treatment with blood plasma from those who had recovered, although a lack of randomized studies makes it hard to assess the treatment's effectiveness.

Perhaps the most striking difference between the two viruses is how long it was before a vaccine was developed. For SARS-CoV-2, it was 10 months from sequencing the virus to producing results from large, randomized trials demonstrating high levels of vaccine efficacy. Large-scale distribution quickly followed. Poliovirus was identified as the pathogen for polio in 1908, but it wasn't until 1955 that US virologist Jonas Salk developed the first effective vaccine to be delivered by means of an injection, followed quickly by an oral vaccine developed by US physician and microbiologist Albert Sabin in 1961.



The next generation of coronavirus vaccines: a graphical guide

Wunsch provides a detailed history of polio, mechanical ventilation, the rise of the field of anaesthesiology, the development of the Salk and Sabin vaccines and the work at Denmark’s Statens Serum Institute, a medical laboratory in Copenhagen, in manufacturing and rolling out the Salk vaccine ten days after it was announced. But she really hits her stride when she describes those whose lives were saved. Another early patient treated by tracheostomy and hand ventilation at the Blegdam hospital was 26-year-old Rosa Abrahamsen. She was a poet, and her beautiful poems, translated into English for the book, begin several chapters.

The Autumn Ghost might have been improved with a timeline, given that it bounces back and forth at many points. Although the extraordinary progress and innovations made in Denmark were central to the development of ICUs, the contribution of parallel efforts from around the world might not have been adequately highlighted.

When I was at the University of Virginia in Charlottesville in the 1970s, I worked as a respiratory technician on the night shift, maintaining Engström

ventilators (alluded to in the book as the “Rolls Royce of artificial respiration”) for patients in the ICU. I had no idea how those ventilators, or indeed ICUs, came to be. But seeing many patients ‘come back to life’ inspired me to go to medical school. Only five decades later, thanks to reading this book, have I learned the remarkable background to these profound innovations — and how the poliovirus radically transformed the future of medicine.

Nature **618**, 234-235 (2023)

doi: https://doi.org/10.1038/d41586-023-01830-3

Updates & Corrections

- **Correction 08 June 2023**: This book review erroneously implied that iron lungs were the direct precursor of modern mechanical ventilators.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01830-3

Opinion

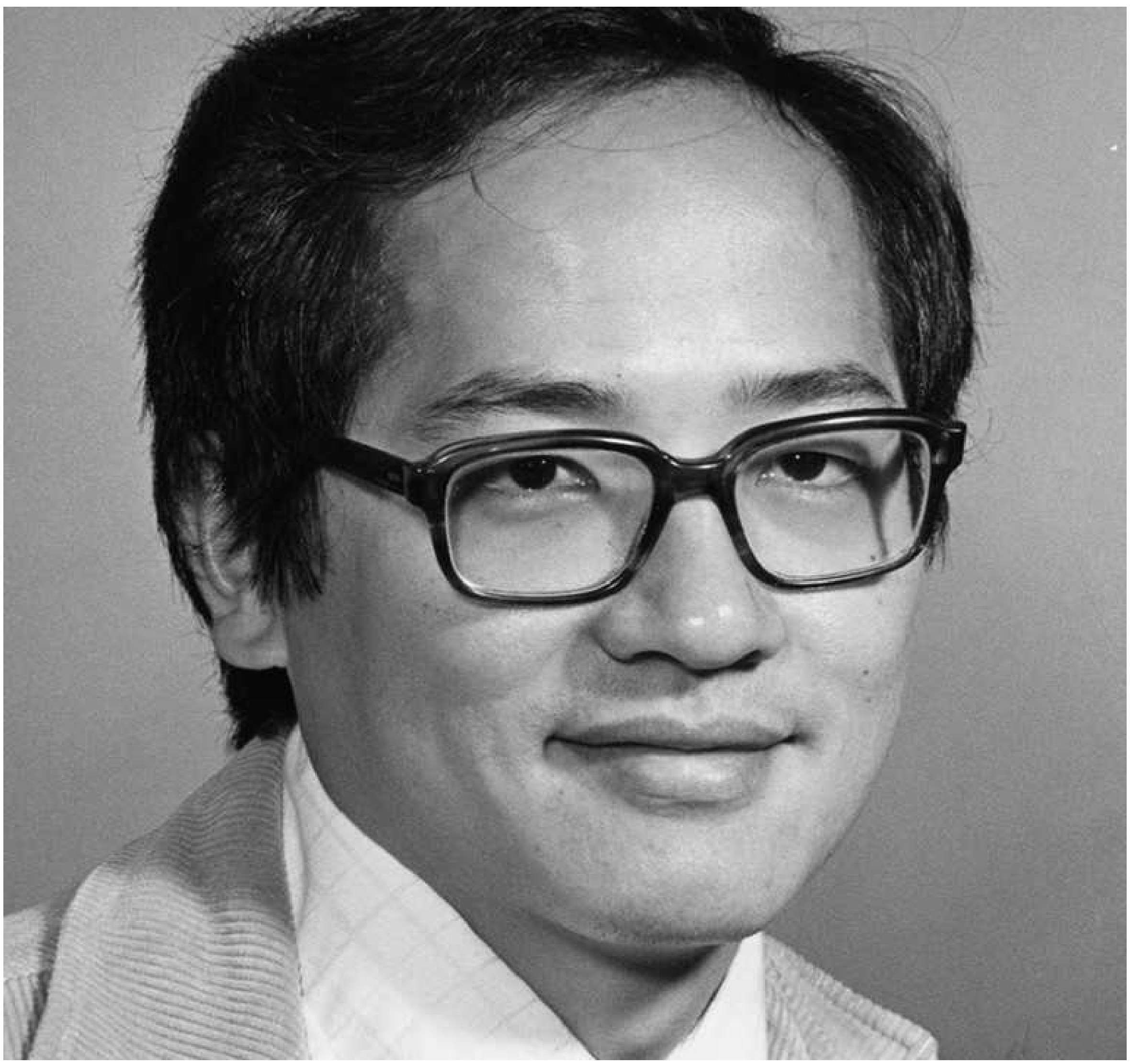
- **Frank Shu (1943–2023)** [05 Jun 2023]
  • Astrophysicist who researched galactic structure and star formation.

- OBITUARY
- 05 June 2023

Frank Shu (1943–2023)

Astrophysicist who researched galactic structure and star formation.

- Douglas Lin &
- Fred Adams



Credit: AIP Emilio Segrè Visual Archives, Physics Today Collection

The Chinese American astrophysicist Frank Hsia-San Shu made seminal contributions to the understanding of galaxies and star formation, leaving a legacy of scientific work characterized by originality and deep insight. He wrote two widely consulted textbooks, and, later in life, embarked on a second career investigating energy technologies to help combat the climate crisis. For his lifelong contributions to theoretical astrophysics, in 2009 he was awarded the Shaw Prize in astronomy — often described as the 'Nobel of the East' — and the Catherine Wolfe Bruce Gold Medal from the Astronomical Society of the Pacific. He died at his home in California on 22 April, aged 79.

Shu was born in 1943 in the southern Chinese city of Kunming, a refuge for intellectuals during the Second World War, owing in part to the evacuation of many universities to the city. He emigrated to the United States at the age of six when his father, the mathematician and engineer Shien-Siu Shu, joined the mathematics faculty at Purdue University in West Lafayette, Indiana. Growing up in the Midwest, he fervently pursued interests in maths and physics. He was admitted to the Massachusetts Institute of Technology in Cambridge, as one of its youngest undergraduate students in physics, and graduated in 1963.

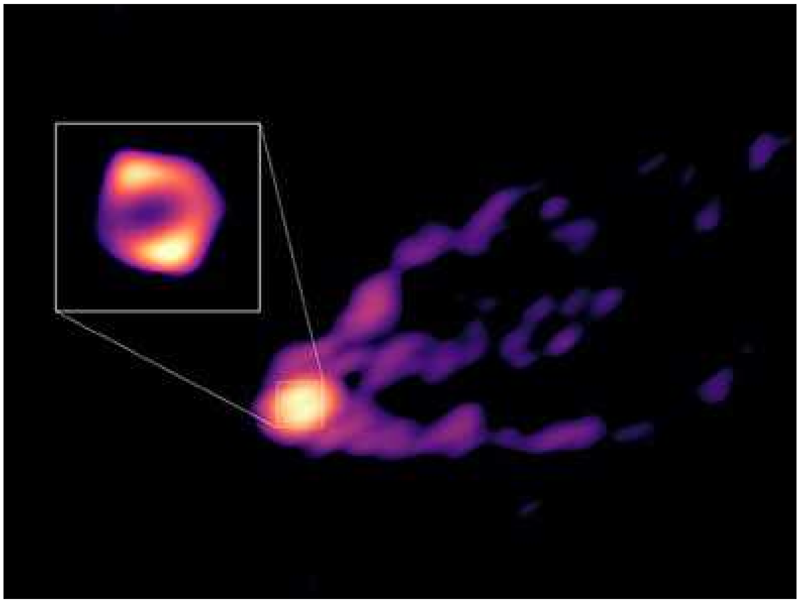


JWST spots planetary building blocks in a surprising galaxy

At MIT, Shu began working with his mentor Chia-Chiao Lin on the dynamics of disk galaxies. This collaboration continued at Harvard University, also in Cambridge, where Shu received his PhD in 1968. Using wave-mechanic methods, Shu and Lin postulated that the spiral arms observed in disk galaxies were a wave phenomenon consisting of spiral patterns propagating through the fluid medium of stars and gas. How these patterns became amplified to their huge size remains a matter of debate, but 'density-wave theory' has allowed predictions of spiral-galaxy properties to be derived from first principles. It has also been used to study the gravitational stability and structure of many astrophysical objects, including galactic disks, binary stars, planet-forming disks orbiting young stars and the rings of Saturn.

Shu's work was motivated by fundamental theoretical issues. He paid special attention to stars not just as points of mass, but as points of light. He was always interested in how they formed, looking at massive OB stars, which light up the spiral structure of disk galaxies viewed at optical and ultraviolet wavelengths. At Stony Brook University in New York City, where Shu moved after his PhD, he led a team that showed how shocks induced by density waves could compress a diffuse stellar medium into a higher-density phase of star-forming giant molecular clouds. In collaboration with radio astronomers at the University of Groningen in the Netherlands, Shu provided further evidence for the density-wave theory.

After joining the faculty at University of California, Berkeley, in 1973, Shu shifted his attention to the formation of individual stars, the subject of his most influential contribution. Although molecular clouds had been identified as the birth sites of new stars, and the general picture of star formation through the collapse of material under its own gravity had been considered, few specific details were known. Shu and his collaborators had to address how nature overcomes the barriers to gravitational collapse, such as angular momentum and ambient magnetic fields.



[Black-hole image reveals details of turmoil around the abyss]

The paradigm that emerged was the now-standard 'four-phase' picture: the formation of cores of higher density in a molecular cloud; their subsequent dynamic collapse; an unexpected (but observable) phase in which strong winds of gas flow out of the still-forming protostar; and the emergence of stars accompanied by planet-forming disks. This framework closely linked theory and observations, and was bolstered further by independent developments, such as the idea that once-radioactive relics found in the most primitive meteorites might originally have been the products of ancient solar flares.

Although some of Shu's theoretical hypotheses remain controversial, he never hesitated to tackle unsettled fundamental concepts and to explore innovative explanations of the most striking observations. He was also a prolific and lucid writer. In addition to many highly cited papers, Shu authored the universally adopted introductory textbook *The Physical Universe* (1982) and the two-volume graduate textbook *The Physics of Astrophysics* (1991–92).

Shu served as the president of the American Astronomical Society from 1994 to 1996. He engaged with a burgeoning astronomical-research community in Taiwan, leading efforts towards the launch of the Institute of Astronomy and Astrophysics, part of the research institute Academia Sinica

based in Taipei. In 2002, he became president of National Tsing Hua University in Hsinchu, Taiwan, following in the footsteps of his father, who had held that position from 1970 to 1975.

Shu returned to the United States in 2006 to the University of California, San Diego, where he became professor emeritus in 2009. For the past dozen years, his concern about the climate crisis led him to study the use of molten-salt reactors to generate energy from nuclear waste and to convert waste biomass into inert products that can be sequestered, removing carbon from the atmosphere. He was also writing an undergraduate textbook called *The Story of Science*, which was to recount how science has underpinned human progress, and the lessons that holds for addressing climate change. That book remains unfinished, but Shu's own place in the story of science is assured.

Nature **618**, 237 (2023)

doi: https://doi.org/10.1038/d41586-023-01831-2

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01831-2

| Section menu | Main menu |

Work

- **When a loved one dies: how academia can support bereaved colleagues** [05 Jun 2023]
  • Finding time and space to grieve for a partner, child or friend can be complicated for researchers — and the peers who support them.
- **Six tips for better coding with ChatGPT** [05 Jun 2023]
  • Although powerful, the tools are not as intelligent as they seem. Use them with caution, computer scientists warn.
- **My mission to grow fruit without the plant** [05 Jun 2023]
  • Lucas van der Zee hopes to restore current farmland to its natural state by eliminating the vegetation stage of crop-growing.

- CAREER FEATURE
- 05 June 2023

When a loved one dies: how academia can support bereaved colleagues

Finding time and space to grieve for a partner, child or friend can be complicated for researchers — and the peers who support them.

- Jyoti Madhusoodanan

Find a new job



Credit: Agnese Abrusci/*Nature*

Organizational psychology researcher Stephanie Gilbert had prepared for her maternity leave months in advance. She compressed her teaching schedule to complete courses in 7 weeks instead of the usual 12 so that she could take the time off. But on her last day of work in 2017, Gilbert experienced the traumatic loss of her 37-week-old baby. When she returned to her job at Cape Breton University in Sydney, Canada, her research, which had focused on how workplace factors contribute to an employee's well-being, took a new turn. Spurred by her personal experience, Gilbert began to study how workplaces support — or fail to look after — grieving employees.

There is little empirical evidence to guide employers or colleagues in helping a person who has experienced the death of a loved one or friend, says Gilbert. But grief is a reality that most adults will face at some point. "It is something that people are dealing with more and more with COVID-19," she says. "But it was always a ubiquitous experience."

However, assistance for people balancing work and grief is often unstructured. Policies around bereavement leave and paid time off can vary widely, often depending on a person's career stage or their relationship to the person who died. The irregular schedules and independent nature of academic work might also make bonds between colleagues more tenuous, Gilbert says.



Science careers and mental health

Grief can affect a person's performance for prolonged periods, and might have different impacts during various career phases. Workplace support for bereaved colleagues must be both proactive and continuous, and needs to go beyond good intentions and condolences, Gilbert says. But often that doesn't happen, in part because of uncertainty about what kinds of help are needed, and because of a lack of clarity about institutional policies.

In recent years, academics have drawn attention to the workplace support needed for people returning to work after having children and those experiencing other life events, says criminal-justice researcher Emily Cooper at the University of Central Lancashire in Preston, UK. "But I don't think academia is very good at talking about grief."

She and others who have experienced the death of a family member or friend while navigating the pressures of academia describe how colleagues could do better by their peers who are in mourning.

Communicate policies clearly

Gilbert's experience of support was relatively straightforward, because plans were already in place for her to be away for several weeks. When her baby died, her institution said she could use her maternity leave to rest and recover both physically and emotionally. But this is not the norm.

The amount of paid bereavement leave a person can take, and whether they can return to work part-time or in a phased manner, depends on individual institutions. Policies can also differ if an individual's child has died aged under 18.

Finding and understanding an institution's policies — and feeling enabled to use them — can seem overwhelming during a time of grief, particularly for early-career researchers who might have shorter contracts or feel less empowered in the workplace than senior faculty members, for instance. "In those early days of grief, it can be very difficult to process and understand what you might need, so there should be ongoing communication," Gilbert says. "And that communication should inform the work accommodations that are offered for grief and bereavement."

In December 2021, Cooper's three-year-old son Alexander had a febrile seizure and died shortly afterwards at the hospital. At the time, Cooper was pregnant and scheduled to be on leave from April 2022. But her employer allowed her to take three months of paid time off starting in December 2021, to be followed by her maternity leave, which enabled her to remain off work until October 2022. "Essentially, I had three months of paid bereavement leave, which is just unheard of in the United Kingdom," Cooper says.

Her manager also offered her the option of returning to work part-time. "I don't think they even put it down as anything official," she says. But knowing she was supported in this way was "a huge relief", she adds. "We could just focus on what we'd gone through, the other children and just surviving without having to worry financially or anything else."



The price of grief

Cooper, who has a permanent contract and is the family's main earner, resumed part-time work in October last year, and returned full time a few months later, in January. But her husband Darren Bowes, who is pursuing a PhD at Lancaster University, UK, had a different experience, and the couple struggled to work out what his funding agency permitted for bereavement leave. "At first, the advice was basically that he should just take holiday or take annual leave," Cooper says. After discussion with the university, Bowes was granted a month off with full pay and additional time if he needed it.

The lack of clarity around doctoral students' bereavement benefits is partly because the policies for staff don't always extend to them, and student policies might apply only to undergraduates. Confusion also emerges when policies cover only specific relationships, as Jenny Fisher, a microbiologist at Indiana University Northwest in Gary, discovered.

When Fisher's long-time partner had a stroke and died in early March 2020, she received messages of condolence and support. When her partner was in hospital, Fisher checked the official family-leave policy online, but it mentioned only leave to care for a child, parent or spouse — and made no allowances for a partner who was not legally a spouse — so she was uncertain whether she could take any time at all. After he died, she enquired about bereavement-leave policies but found there weren't any explicitly written. "I wish there had been clearer policies," she says.

Precarious positions

For Fisher, who was not tenured at the time, the COVID-19 pandemic compounded her uncertainty. Immediately after the week of spring break — the only bereavement leave that Fisher took — her university shut down. Fisher resumed work and moved the three undergraduate courses she was teaching to an asynchronous online format. Although there was no explicit pressure to keep up her work, and she doesn't doubt that her department would have supported her, Fisher struggled to ask for help. "I don't know if I didn't feel like I could ask for help," she says. "I don't know how much of it was just my own need to get through, to be able to do things and seem competent."



When Amanda Moehring’s spouse died in 2020, her university colleagues helped to cover both her professional and personal tasks, including extra supervision of graduate students and grocery deliveries.Credit: Theresa Rice

This pressure to perform can be especially acute for early-career researchers, graduate students and postdoctoral researchers who, in addition to dealing with personal grief, are faced with professional concerns such as short contracts and the need to produce high-quality work quickly, Gilbert says. That pressure is reinforced by academia’s standards: in their work, Gilbert and her colleagues have spoken with pre-tenure faculty members who have struggled to balance the need to take time off for bereavement with expectations about research productivity. Time off can have real repercussions for researchers’ progress, she says, and one person described the negative feedback he received during tenure review as “a kick in the pants from his department after what he’d been through”.

For Fisher, the pandemic led to an automatic extension on the tenure clock, or time window before applying for tenure, but she did not take any extra time off to cope with her loss, in part because she felt a responsibility towards her students. In retrospect, Fisher wishes she had received more proactive support to cope with her responsibilities, and for university policies to be framed with an eye towards “how to give people a break without ruining the experience for the students”, she says.

Cooper remembers the “insecurity and pressure” she felt in the pre-tenure stages of her career, and stresses the need for institutions to be particularly supportive of early-career researchers coping with grief. If she had experienced the loss of her son “while in a precarious job”, she says, “it’s very possible I wouldn’t have stayed in academia”.

Surrounded by support

Assistance from colleagues and institutions is especially meaningful when it is anticipatory and tailored to an individual’s needs. When geneticist Amanda Moehring’s spouse died suddenly in March 2020, her institution — the University of Western Ontario in London, Canada — provided her with proactive and ongoing help despite the pandemic. “The thing that I think my university did particularly well is anticipatory support,” she says. “I didn’t have to ask for anything, aside from telling them that my husband died and I had to go on leave.”

As soon as she did, her colleagues jumped in to help her in her work and personal life, including arranging grocery deliveries and tutors for her children during online school. “It saved me the burden of having to think of what I needed,” Moehring says.

For five months, her colleagues redistributed Moehring’s service and teaching responsibilities to other staff members, and arranged for her graduate students to receive counselling and to be paired with alternative supervisors.

The institution also confirmed that Moehring felt comfortable with the plans, which was as important as providing help. “They took care of things, but

they also didn't do it behind my back," she says. "I didn't have to think of coordinating it, but I still felt like I had some say in what was happening."



Claire Carmalt (left) received care from colleagues when her daughter Tory died in 2022.Credit: Claire J. Carmalt

Those in more senior roles also benefit from proactive support. When organic chemist Claire Carmalt's daughter died in October 2022, her department at University College London rallied support for her graduate students and to relieve her teaching load. For Carmalt's extra responsibilities as department head, her institution set up an alternative e-mail address to delegate tasks to others. Carmalt was informed of these changes and let the university know when she was ready to receive those e-mails again. At that point, "the messages of support were actually fantastic", she says.

Returning to work

Despite her senior role, returning to work was not easy for Carmalt, because, like many in mourning, she was uncertain of her own reactions or how people would react to her presence. “I’ve been there 25 years, but I felt very anxious going back,” she says. “Would I get upset all the time? Would I be able to keep it together in meetings? Make decisions, provide scientific guidance and all of that?”

The accommodations researchers might need when they resume work will vary. Some might prefer to return part-time at first; others might only be able to cope with a subset of tasks or need a private workspace. Communicating about work accommodations is crucial and should be an ongoing discussion, Gilbert says. The deep thinking required to write manuscripts or glean meaning from data can be especially challenging when grieving, she adds, so faculty members and students might prefer to shift their focus towards less intensive tasks, such as data collection.



The price of grief

Teaching a large undergraduate class can seem especially daunting, and many researchers might feel more comfortable resuming mentoring responsibilities and teaching smaller, graduate-level courses. “There’s a

great deal of emotional labour required when teaching and dealing with students, and having to maintain that professional appearance," Gilbert says. "For some people, that can be quite a demand."

Cooper, for instance, found she isn't ready to teach yet. "I'm just not in a position to go into a room and motivate a group of students," she says. She was able to discuss with her supervisor which responsibilities she would resume when she returned to work. "I've found it really helpful to still have a voice in the process," she says.

Department heads and other colleagues can also ease a person's transition back to work by simply checking in with them about how they might like to be treated on their return and letting other colleagues know, Moehring says. This could be as simple as asking whether a person would like colleagues to acknowledge their loss in the workplace or stick to work topics only, because individual preferences can vary widely.

Having a safe physical space, such as a private office or quiet room, can be helpful because grief can manifest differently each day, Carmalt says. "The problem with grief is, I probably look like I'm fine today," she says. "But yesterday I was struggling at work. Out of the blue, for no reason, a big wave of grief will come, and trying to fight back those tears can be very difficult."

It's also important for managers to keep an open mind about work schedules for several months, particularly because grief can disrupt people's sleep schedules. Although those in senior roles similar to Carmalt's can choose to work from home, that might not be the case for postdocs or graduate students working on experiments, she says. "Supervisors need to have some flexibility with understanding that their days are going to be different."

When a colleague dies

Grief in the workplace is amplified when the person who has died is a colleague.

When Peter Salvino, a graduate student in neuroscientist Lucas Pinto's laboratory at Northwestern University in Evanston, Illinois, died suddenly in

December 2022, Pinto's department chair told him that he could step away from his teaching responsibilities and committee duties if he needed to. "That was very helpful, and I did take them up on it," Pinto says.

Alongside his own grief, Pinto found himself needing to support other students and lab members who looked to him for guidance as principal investigator (PI). He spoke to each group member individually to break the initial news and clarified that the lab would support their needs. He gave students the option of working when and where they chose, and coordinated with the university to ensure that counselling services were available to them over the winter break. Because the students were good friends with Salvino, they are motivated to continue his work while coping with grief themselves. "People rally around the desire to see his work through, see his name on publications," Pinto says. But at the same time, "we're trying to do our best to just keep going".

Although the trauma didn't change Pinto's approach to mentoring, he has noticed a shift in his interactions with lab members. "I'm always friendly, but this kind of gave me a licence that I didn't know I needed, to really be myself without the PI sheen," he says. "It brought into really sharp focus how important it is to have a healthy, welcoming lab environment. Good science follows from that."

Whether at the level of an individual lab or the broader academic community, solidarity through periods of grief can have lasting benefits for everyone. Supporting employees and colleagues through a traumatic period is likely to pay off in terms of retention and positive job attitudes, Gilbert says.

But, more importantly, she adds: "People deserve to be cared for by their workplace in that vulnerable time."

Nature **618**, 419-421 (2023)

doi: https://doi.org/10.1038/d41586-023-01832-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01832-1

- TECHNOLOGY FEATURE
- 05 June 2023

Six tips for better coding with ChatGPT

Although powerful, the tools are not as intelligent as they seem. Use them with caution, computer scientists warn.

- Jeffrey M. Perkel

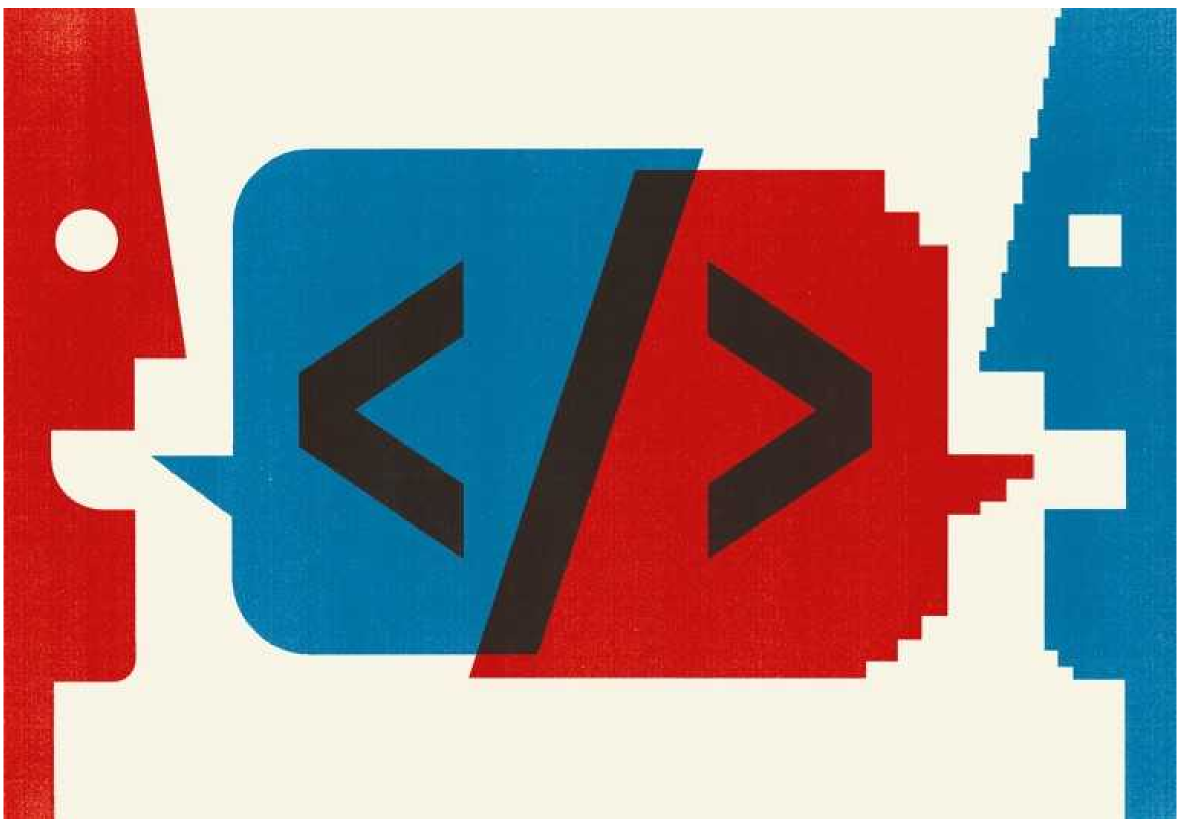


Illustration by The Project Twins

Unless you've been living under a rock, you know about ChatGPT. The chatbot, driven by artificial intelligence (AI) and created by OpenAI in San

Francisco, California, provides eerily human-like responses to user questions (called prompts) on almost any subject. ChatGPT is trained on a vast corpus of text, and its ability to engage in text-based conversation means that users can refine its responses. Even if its initial answers are wonky, it often eventually produces accurate results, including software code.

Researchers can use ChatGPT to debug and annotate code, translate software from one programming language to another and perform rote, boilerplate operations, such as plotting data. A March preprint reported that the program could solve 76% of 184 tasks in an introductory bioinformatics course, such as working with spreadsheets, after a single try and 97% within seven attempts[1].

That's good news for researchers who feel uncomfortable coding, or who lack the budget to employ a full-time programmer — for them, chatbots can be a democratizing tool.

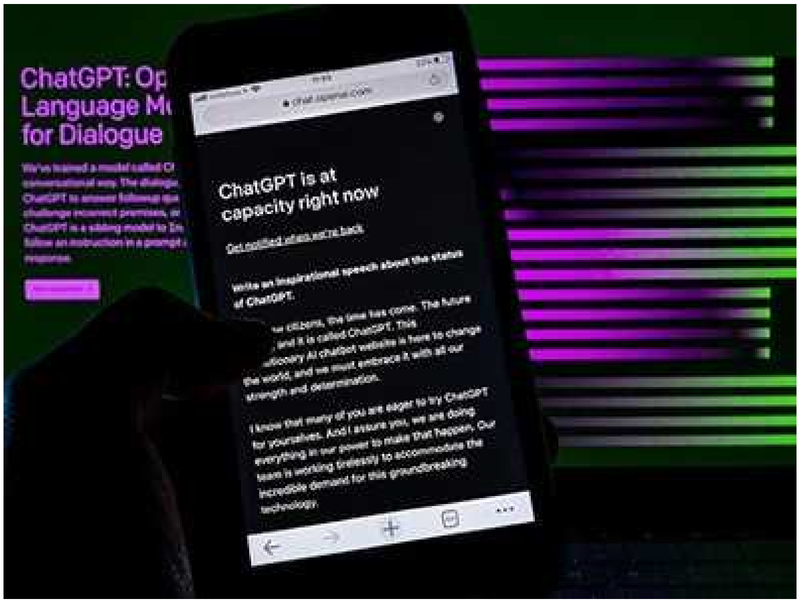
Yet for all their apparent sentience, chatbots are not intelligent. They have been called stochastic parrots, randomly echoing back what they've seen before. Amy Ko, a computer scientist at the University of Washington in Seattle, invokes a long-running US quiz show to describe the tool's limitations, writing on the Mastodon social-media site: "ChatGPT is like a desperate former Jeopardy contestant who stopped following pop culture in 2021 but really wants to get back into the game, and is also a robot with no consciousness, agency, morality, embodied cognition, or emotional inner life." (The data used to train ChatGPT only extend into 2021.)

In short, ChatGPT and related tools based on large language models (LLMs), which include Microsoft Bing and GitHub Copilot, are incredibly powerful programming aids, but must be used with caution. Here are six ways to do so.

Choose your applications

Chatbots work best for small, discrete programming tasks, such as loading data, performing basic data manipulations and creating visualizations and

websites. But that’s not the same as software engineering, says Neil Ernst, a computer scientist at the University of Victoria in Canada.



[ChatGPT: five priorities for research]

“Software engineering is a lot more than just solving a programming puzzle,” Ernst explains. “There’s thinking about test frameworks, writing maintainable code and understanding the trade-offs involved in building a system” — such as that between speed and readability. “I don’t think that the current tools are solving any of those issues.”

That leaves many tasks they can do, such as remembering the syntax for creating visualizations with Matplotlib, a graphing library for the programming language Python. In that sense, chatbots are like a conversational interface to Stack Overflow, an online question and answer forum for programmers. “That’s not stuff that anyone particularly enjoys writing,” says Ernst, “and it saves time for us to ask the tough analytical questions that we might have about the data.”

Chatbots are also good at explaining why code doesn’t work. Emery Berger, a computer scientist at the University of Massachusetts Amherst, has exploited those abilities to build several helpful tools. One, called cwhy, uses ChatGPT to explain compiler errors in code written in the programming languages C, C++ and Rust. Another, ChatDBG, provides a conversational

interface for debugging, and a third, Scalene, uses the AI to suggest code optimizations to improve performance.

Chatbots can even translate code from one programming language to another. Mathieu Coppey, a biophysicist at the Curie Institute in Paris, is using ChatGPT to help him move from MATLAB, his preferred language, to Python. Using Google and online forums, he typically needs days to get his Python code working. “Now, I can do that in an hour or so,” he says.

Trust, but verify

Chatbots might not always know what they’re talking about, but they certainly sound like they do. In some cases, the AI doesn’t understand the question; at other times, it provides an incorrect answer. When the code fails to run, such mistakes are obvious. Sometimes, however, the code runs but yields the wrong result.

According to a study[2] co-authored by linguist Emily Morgan at the University of California, Davis, chatbots — like the human-written code on which they were trained — often create what she calls “simple, stupid bugs”. These single-line errors, such as using > instead of >= in a conditional statement, are easy to fix, but hard to find. “If you don’t know enough to tell the difference between something correct and something that’s effectively nonsense, then you could get yourself in trouble,” she says.



What ChatGPT and generative AI mean for science

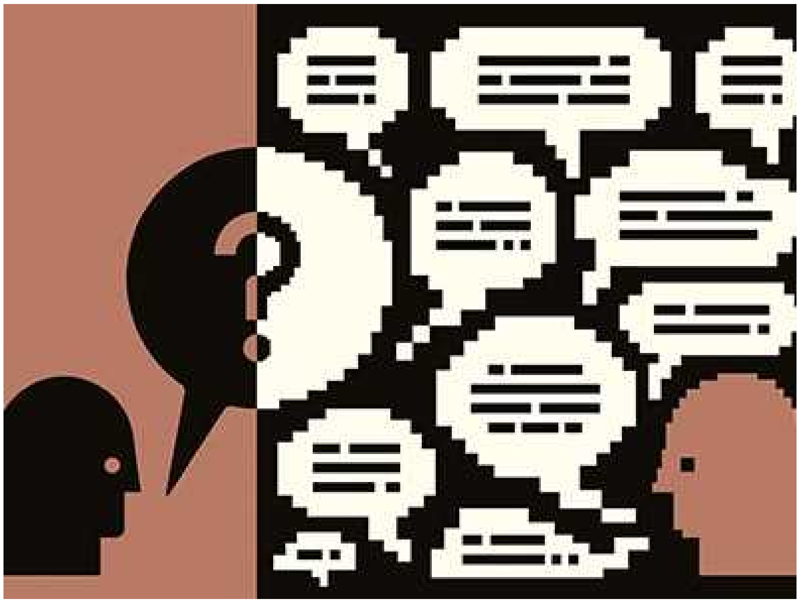
Iza Romanowska, a complexity scientist who studies ancient civilizations at the Aarhus Institute of Advanced Studies in Denmark, has used ChatGPT to produce code in a language called NetLogo. Because there's less online code written in NetLogo than in the languages Python and R, ChatGPT is less fluent in it. Sometimes, the AI peppers its suggested code with functions that don't actually exist, she says — a behaviour sometimes called hallucination.

The bottom line is not to blindly accept what ChatGPT gives you — read it carefully and test it. Make sure it performs as expected on 'edge cases' — for instance, does an algorithm to sort *n* numbers include the *n*th number? Patrick Lam, a computer scientist at the University of Waterloo in Canada, says: "I wouldn't trust this further than I can throw it."

Think safety

Chatbots output code that reflects their training data. That's not always a good thing, says Ko. "The aggregate quality of code on the web that's shared, that these [chatbots] are trained on, is actually quite low."

Just as random code online is unlikely to be particularly efficient or robust, so too is chatbot-generated code. It might not work well on large data sets, for instance, and can contain security vulnerabilities.



[Could AI help you to write your next paper?]

Brendan Dolan-Gavitt, a computer scientist at New York University, says that when Github's Copilot programming tool launched in 2021, he and his team tested it in 89 security-relevant scenarios. One was the ability to check for malformed queries using the language SQL that could corrupt a database — known as an SQL-injection attack[3]. "About 40% of the time, Copilot was producing code that was vulnerable." That's a moving target — when Dolan-Gavitt put those scenarios to a newer version of the LLM underlying ChatGPT, called GPT-4, the error rate fell to 5%.

Still, it pays to check your code. But also consider the application — not everything is mission-critical. The web interface to a database or visualization tool, for instance, might require extra vigilance. But if you know what the answer to your programming problem should look like, "Just go for it", says computer scientist Sayash Kapoor at Princeton University in New Jersey, "because it's easy to check if you're wrong."

Iterate

Chatbot-based coding, says Ko, “is not a single-shot sort of experience”. It’s a conversation. “You write something, you get something back, you read it sceptically, you ask for more detail, you ask for it to fix something.”

Gangqing (Michael) Hu, who runs the bioinformatics core facility at West Virginia University in Morgantown, capitalized on that iterative workflow to develop a method that beginners in bioinformatics can use to optimize chatbot prompts, called OPTIMAL[4]. Users provide detailed prompts, test the replies and feed back to the chatbot to tweak its responses. That can include questions about errors as well as tweaks to the prompt itself. “Communication is the key,” Hu explains.

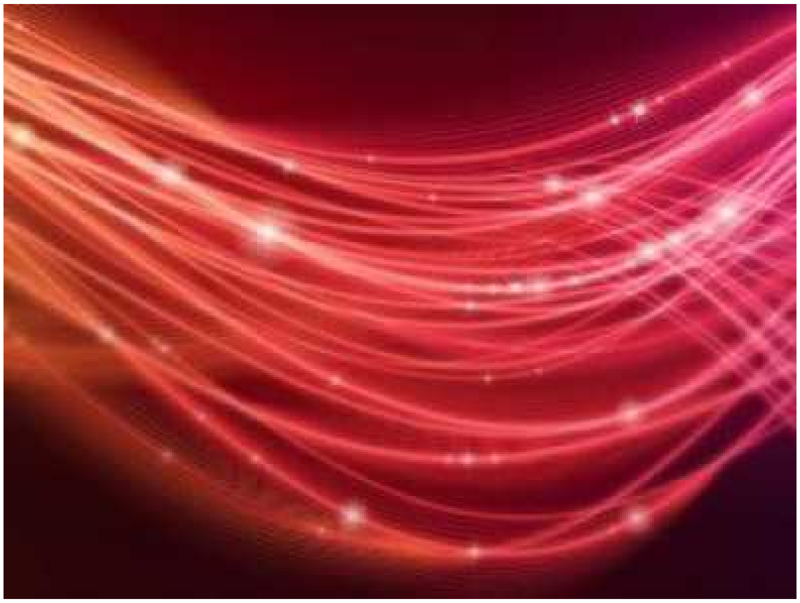
If you get stuck, try adjusting the settings, suggests Xijin Ge, a bioinformatician at South Dakota State University in Brookings. The ChatGPT ‘temperature’ setting, for instance, controls creativity — the higher the temperature, the more creative the output. “Sometimes it works,” Ge says.

But not always — in some cases, “you’ll have to intervene and take over”, says Ko.

Anthropomorphize

Chatbots aren’t people, but it can be helpful to treat them that way. “Treat this AI as a summer intern,” Ge advises — a college student who is hard-working and eager to please, but also inexperienced and error-prone.

Avoid ambiguity, and break your problem up into smaller pieces, suggests Paul Denny, a computer scientist at the University of Auckland, New Zealand.



NatureTech hub

Another tip: direct the chatbot to assume a role, such as a biologist who is fluent in Python. Specify the tools or programming libraries you would like to use. Such directives can help the chatbot to get “into the right probabilistic space”, says Ko — that is, home in on the text that is most likely to follow the prompt.

For instance, one prompt in Hu’s study[4] asked ChatGPT: “Act as an experienced bioinformatician proficient in ChIP-Seq data analysis, you will assist me by writing code with number of lines as minimal as possible. Reset the thread if asked to. Reply “YES” if understand.”

And, if possible, provide starting code, comments and expected results. “Examples can really help ChatGPT to target it in the right direction,” says Dong Xu, a computer scientist at the University of Missouri, Columbia.

Embrace change

Finally, LLMs are constantly evolving, and becoming more powerful. That’s good news for researchers, although it will keep them on their toes. Prompt lengths are increasing, allowing for more nuanced responses. And new tools are constantly emerging. One plug-in called Code Interpreter turns ChatGPT

into a digital data analyst, allowing users to upload data sets, ask questions of their data and download results. As one blogger on AI put it, “It’s like having a conversation with your data. How cool is that?”

Nature **618**, 422-423 (2023)

doi: https://doi.org/10.1038/d41586-023-01833-0

References

1. Piccolo, S. R. *et al.* Preprint at https://arxiv.org/abs/2303.13528 (2023).
2. Jesse, K. *et al.* Preprint at https://arxiv.org/abs/2303.11455 (2023).
3. Pearce, H. *et al.* Preprint at https://arxiv.org/abs/2108.09293 (2021).
4. Shue, E. *et al. Quant. Biol*. https://doi.org/10.15302/J-QB-023-0327 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01833-0

- WHERE I WORK
- 05 June 2023
- Correction 06 June 2023

My mission to grow fruit without the plant

Lucas van der Zee hopes to restore current farmland to its natural state by eliminating the vegetation stage of crop-growing.

- James Mitchell Crow



Lucas van der Zee is a horticulture and product physiologist at Wageningen University, the Netherlands. Credit: Francesco Rucci and Francesco Marinelli for *Nature*

Farming today is unsustainable, with its greenhouse-gas emissions and its destruction of wildlife habitats. Agriculture now takes up half of Earth's hospitable land surface, at the expense of these habitats.

Developing sustainable food production will require multiple approaches. My PhD research at Wageningen University in the Netherlands is inspired by alternative, indoor methods of food production, such as farming leafy green vegetables vertically and growing meat in a laboratory. I wondered — if you can grow meat without an animal, can you grow fruit without a plant?

This could be done inside, which might help us to return some agricultural land to nature.

During my master's programme in agricultural engineering, I discussed indoor fruit production with my professor. He invited me to write a master's thesis on the subject of growing fruits without the plant. Now, we're testing this idea.

The main task of my PhD is to collect immature fruit or flowers from tomato plants, and try to grow them into healthy, high-quality tomatoes in the lab. Ultimately, we want to produce fruit from tomato meristem cells — undifferentiated cells from which new plant organs can grow — and skip using a plant entirely.

The jars in this photo contain some of the early results of my research, when I was testing different growing conditions at different stages of tomato development. The lighting is an energy-efficiency measure; we use blue and red LEDs, the best colours for photosynthesis.

Some people get excited about the possibilities of my work, but most are hesitant, which I understand. In the past, people have rushed into using food technologies, such as cage-grown chickens and the extensive use of synthetic fertilizer, without considering the potential downsides. An important part of my PhD programme will be to assess the sustainability of my method.

Nature **618**, 426 (2023)

doi: https://doi.org/10.1038/d41586-023-01834-z

Updates & Corrections

- **Correction 06 June 2023**: This article erroneously stated that Lucas van der Zee is trying to grow stems into high-quality tomatoes. In fact, he wants to grow tomatoes from immature fruit and flowers.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01834-z

Research

- ## Interchange reconnection as the source of the fast solar wind within coronal holes [07 Jun 2023]
  • Measurements of fast solar wind streams from the Parker Solar Probe spacecraft provide strong evidence for the interchange reconnection mechanism being responsible for accelerating the fast solar wind.
- ## Faster sorting algorithms discovered using deep reinforcement learning [07 Jun 2023]
  • Artificial intelligence goes beyond the current state of the art by discovering unknown, faster sorting algorithms as a single-player game using a deep reinforcement learning agent. These algorithms are now used in the standard C++ sort library.
- ## Non-Abelian braiding of graph vertices in a superconducting processor [11 May 2023]
  • A unitary protocol for braiding projective non-Abelian Ising anyons in a generalized stabilizer code is implemented on a superconducting processor, allowing for verification of their fusion rules and realization of their exchange statistics.
- ## Dynamic optical response of solids following 1-fs-scale photoinjection [24 May 2023]
  • Petahertz-scale optical-field metrology in a pump-probe setting enables the direct observation of how the optical properties of a medium evolve after 1-fs-scale photoinjection.
- ## X-ray-to-visible light-field detection through pixelated colour conversion [10 May 2023]
  • Lithographically patterned perovskite nanocrystal arrays were used to determine radiation vectors from X-rays to visible light and the emission colours of the nanoparticles was used to create images of three-dimensional objects and for phase-contrast imaging.
- ## Tree islands enhance biodiversity and functioning in oil palm landscapes [24 May 2023]
  • A large-scale, five-year study in Indonesia finds that enriching oil palm-dominated landscapes with patches of trees bolsters biodiversity and ecosystem functioning without impairing oil palm yields but should not replace forest protection.
- ## Ancient human DNA recovered from a Palaeolithic pendant [03 May 2023]

• A non-destructive DNA isolation method for the stepwise release of DNA trapped in ancient tooth and bone artefacts is developed.

- **Pan-cancer whole-genome comparison of primary and metastatic solid tumours** [10 May 2023]
  • The genomic differences between primary and metastatic tumours are assessed across 23 cancer types using pan-cancer whole-genome analysis.
- **Myelin dysfunction drives amyloid-β deposition in models of Alzheimer's disease** [31 May 2023]
  • Mouse models show that myelin dysfunction and associated inflammation increase with age, which can promote amyloid-β deposition and therefore risk of developing Alzheimer's disease.
- **A Pseudomonas aeruginosa small RNA regulates chronic and acute infection** [24 May 2023]
  • A study examining bacterial gene expression in human-derived samples identifies a gene encoding a small RNA and describes how it orchestrates the transition between chronic and acute infection in Pseudomonas aeruginosa.
- **Deterministic evolution and stringent selection during preneoplasia** [31 May 2023]
  • We model occult preneoplasia by biallelic inactivation of TP53, a common early event in gastric cancer, in human gastric organoids, the results implying predictability in the earliest stages of tumorigenesis.
- **Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum** [24 May 2023]
  • Ubiquitination of the receptor FAM134B regulates ER-phagy and remodelling of the endoplasmic reticulum in response to cellular demands.
- **Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy** [24 May 2023]
  • The membrane-shaping protein ARL6IP1 is involved in the selective degradation of the endoplasmic reticulum, and this process depends on its ubiquitination and interaction with other membrane-shaping proteins such as FAM134B.

- Article
- Open Access
- Published: 07 June 2023

Interchange reconnection as the source of the fast solar wind within coronal holes

- S. D. Bale ,
- J. F. Drake,
- M. D. McManus,
- M. I. Desai,
- S. T. Badman,
- D. E. Larson,
- M. Swisdak,
- T. S. Horbury,
- N. E. Raouafi,
- T. Phan,
- M. Velli,
- D. J. McComas,
- C. M. S. Cohen,
- D. Mitchell,
- O. Panasenco &
- …
- J. C. Kasper

Show authors

Nature **volume 618**, pages 252–256 (2023)

- 4337 Accesses

- 613 Altmetric
- Metrics details

Abstract

The fast solar wind that fills the heliosphere originates from deep within regions of open magnetic field on the Sun called 'coronal holes'. The energy source responsible for accelerating the plasma is widely debated; however, there is evidence that it is ultimately magnetic in nature, with candidate mechanisms including wave heating[1,2] and interchange reconnection[3,4,5]. The coronal magnetic field near the solar surface is structured on scales associated with 'supergranulation' convection cells, whereby descending flows create intense fields. The energy density in these 'network' magnetic field bundles is a candidate energy source for the wind. Here we report measurements of fast solar wind streams from the Parker Solar Probe (PSP) spacecraft[6] that provide strong evidence for the interchange reconnection mechanism. We show that the supergranulation structure at the coronal base remains imprinted in the near-Sun solar wind, resulting in asymmetric patches of magnetic 'switchbacks'[7,8] and bursty wind streams with power-law-like energetic ion spectra to beyond 100 keV. Computer simulations of interchange reconnection support key features of the observations, including the ion spectra. Important characteristics of interchange reconnection in the low corona are inferred from the data, including that the reconnection is collisionless and that the energy release rate is sufficient to power the fast wind. In this scenario, magnetic reconnection is continuous and the wind is driven by both the resulting plasma pressure and the radial Alfvénic flow bursts.

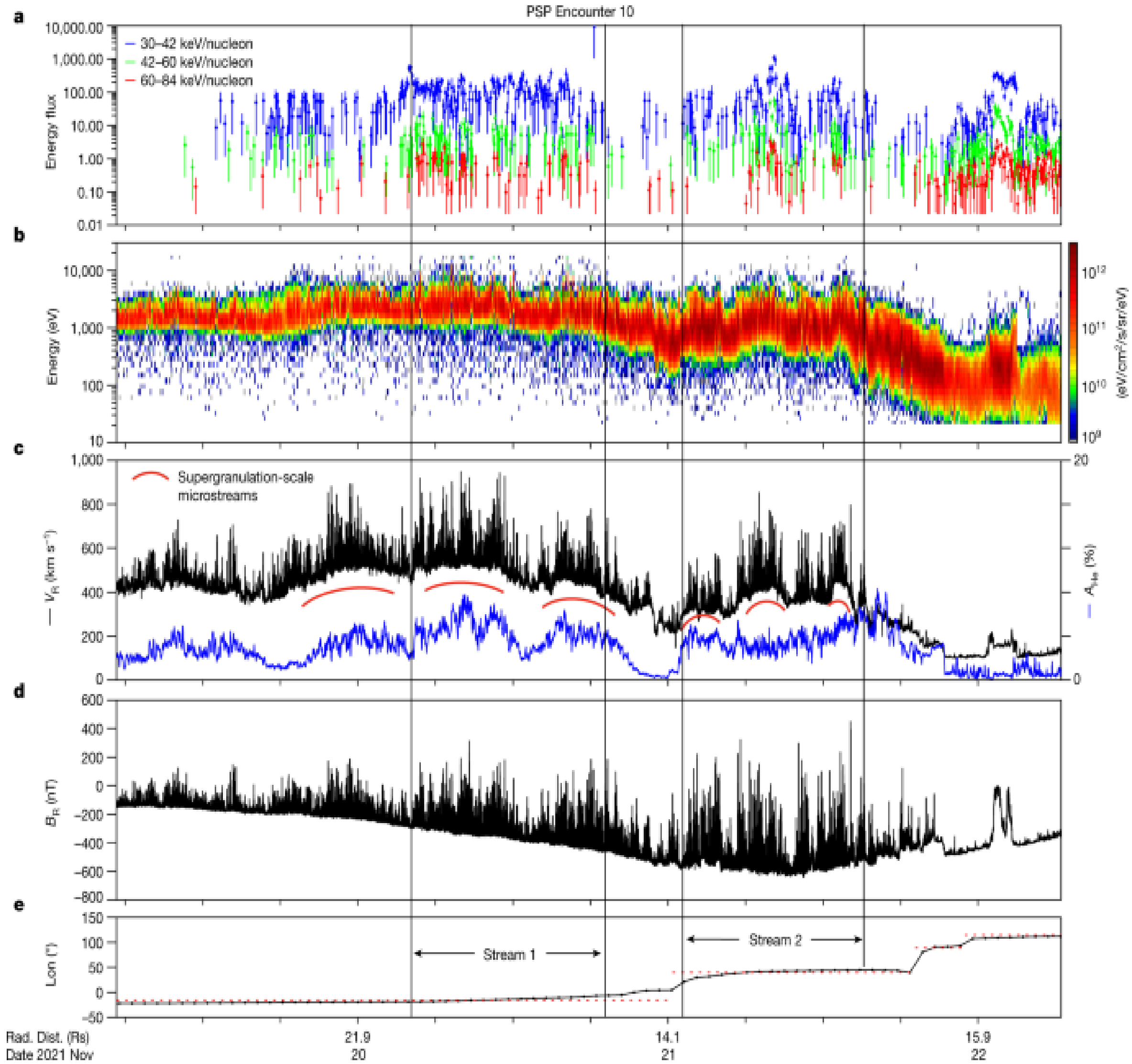
Main

Recent measurements from the NASA Parker Solar Probe (PSP) showed that the solar wind emerging from coronal holes is organized into 'microstreams' with an angular scale (5–10°) in the Carrington longitude[9] similar to the underlying supergranulation cells associated with horizontal flows in the photosphere[10]. However, the footpoints of the previous PSP encounter were

at high latitudes on the far side of the Sun, so that the magnetic structure of the cells and their connectivity to the spacecraft could not be determined, preventing a complete analysis of the source of the microstreams.

On solar Encounter 10 (E10), PSP came within 12.3 solar radii (R_S) of the photosphere. Figure 1 summarizes the plasma[11], energetic ion[12] and magnetic field measurements[13] made near perihelion. An ion spectrogram in Fig. 1a,b extends from thermal energies to around 85 keV and, like the proton velocity in Fig. 1c, is structured as discrete 'microstreams'[9,14,15] whose duration decreases from around 10 h to around 2 h as the spacecraft approaches perihelion. Data in Fig. 4b (and discussed later) show that the ion energy distributions are power laws at high energy that extend to beyond 100 keV. The characteristic structure of the microstreams is highlighted by red arcs in Fig. 1c, and a blue trace indicates the measured thermal alpha particle abundance $A_{He} = n_\alpha/n_p$ (where n_α and n_p are the alpha particle density and proton number density, respectively), which is similarly modulated. The high first ionization potential of helium requires that the alpha particle abundance is frozen-in at the base of the corona or in the chromosphere[16], so these microstream structures are organized at the source of the wind itself. The radial component of the interplanetary magnetic field in Fig. 1d shows that large-amplitude, Alfvénic field reversals, 'switchbacks', are also associated with the microstreams. A potential field source surface (PFSS) model[17,18,19] (Methods) is used to infer the footpoints of the magnetic field that connects to the PSP and shows connection to two distinct coronal holes. The time series of the longitude of the footpoint on the solar surface is shown in Fig. 1e and as white diamonds against a 193-Å Solar Dynamics Observatory/Extreme Ultraviolet[20] image in Fig. 2a.

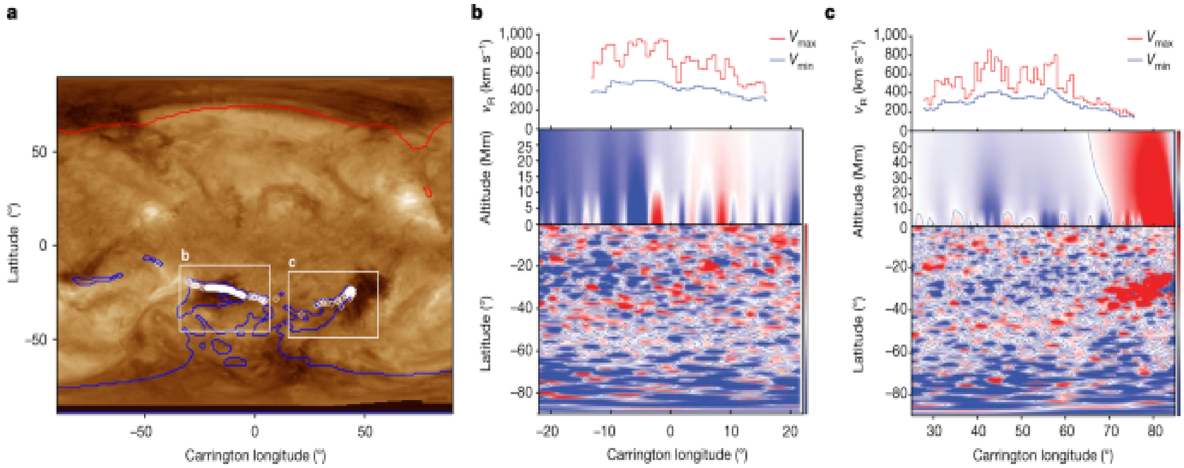
Fig. 1: Time series measurements of the solar wind plasma and magnetic field through the November 2021 solar encounter.



a,**b**, Hot solar wind ions in **a** extend in energy to greater than 85 keV as suprathermal tails on the proton particle distribution in **b**. **c**, Red arcs mark the solar wind radial velocity (V_R) microstream structure that is organized in Carrington longitude at angular scales associated with supergranulation convection and the photospheric network magnetic field (Fig. 2). These microstreams become shorter in duration as the spacecraft accelerates through perihelion near the centre of this figure and sweeps more rapidly through Carrington longitude. The thermal alpha particle abundance (A_{He}, blue trace in **c**) is similarly modulated by the microstream structure. The alpha particle abundance is frozen-in at the base of the corona. **d**, Reversals of the radial magnetic field (B_R), so-called 'switchbacks', are organized by the microstreams and are linked to the radial flow bursts by the

Alfvénicity condition. **e**, Photospheric footpoints from a PFSS model instantiation indicate two distinct coronal hole sources well separated in Carrington longitude (Lon), shown in Fig. 2 (and as dotted lines in Fig. 1e).

Fig. 2: Solar wind during PSP Encounter 10 emerges from two coronal holes.

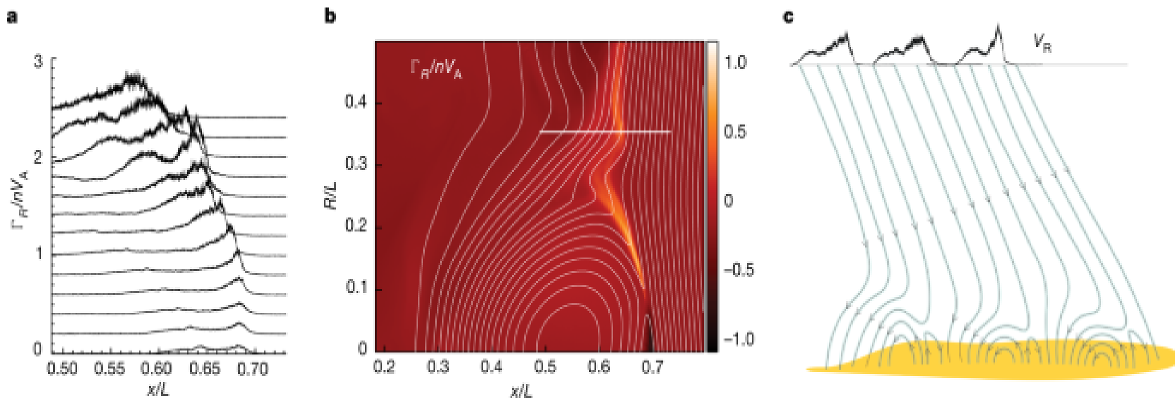


a, An extreme ultraviolet (193 Å) map of the corona shows cooler regions (darker pixels) associated with open magnetic field within two separate, near-equatorial coronal holes. A PFSS model maps the interplanetary magnetic field from the PSP spacecraft to footpoints (white diamonds) within the coronal holes. **b**, The magnetic field and velocity microstream profile within the first coronal hole: the upper panel shows minimum (blue) and maximum (red) radial speed versus longitude, and the second panel shows the vertical magnetic field along the footpoints extending from the photosphere to 30 Mm from magnetogram measurements and a PFSS model that accounts for the motion of the spacecraft. The bottom panel is a map of the magnetic field polarity just above the photosphere, again from the PFSS model. **c**, The corresponding structure within the second coronal hole. These data indicate that the radial magnetic field is organized into mixed radial polarity intervals on the same scales as the velocity microstreams observed by PSP.

Correspondence of the temporal structure of the switchback and radial velocity bursts with the spatial periodicity of the surface magnetic field documented in Figs. 1 and 2 suggests that magnetic reconnection between open and closed magnetic fields in the low corona (interchange reconnection) is the driver of these bursts[9,21,22,23]. Interchange reconnection

in the weakly collisional corona is expected to be bursty rather than steady[24,25,26,27]. The energetic ions and enhanced pressure in these bursts are also signatures of reconnection[28,29,30]. The data suggest that it is a continuous process in the source regions of open flux. Figure 3c is a schematic that shows open flux reconnecting with closed flux regions in the low corona. In this figure the open flux migrates to the left, reconnecting with successive regions of closed flux, with the consequence that the bursty outflow from interchange reconnection fills all of the open flux, as seen in the data.

Fig. 3: Schematic of interchange reconnection and the structure of the reconnection exhaust from a simulation.



a, A time sequence of the dimensionless radial flux along the cut shown by the horizontal white line in **b**. Each successive cut is separated by a time 0.037 L/V_A, where L/V_A is the Alfvén transit time across the simulation domain, and is shifted upward to avoid overlapping the data. The cuts reveal the bursty nature of the outflow resulting from the generation of flux ropes within the elongated current layer[24,25,26,27]. As shown in the schematic, newly reconnected field lines have higher outflow fluxes than field lines reconnected earlier in time. **b**, The dimensionless radial flux with overlying magnetic fields in white from a PIC simulation of interchange reconnection showing Alfvénic upward and downward flows from the reconnection site above the coronal surface. Details about the simulation set-up are found in the Supplementary material. The reconnected magnetic field migrates to the left as it straightens and drives the outflow exhaust. **c**, A schematic of reconnection between open and closed magnetic flux (interchange reconnection) in the low corona based on the PSP data shown in Fig. 1. The data suggest that reconnection between open and closed flux is nearly

continuous. In the schematic, the open magnetic flux is moving continuously to the left. An open field line first reconnects with the closed flux above the solar surface, forming upward and downward oriented loops. The open bent field then straightens and drives Alfvénic flow outward. As it moves to the left, the open field line then intersects another closed flux region and the process repeats. Thus, the open flux is completely filled with high-speed outflowing plasma – the exhaust from interchange reconnection. The cuts of the radial velocity measured by an observer crossing the open flux at the top of the schematic indicate that the highest-speed bursty outflows are on newly reconnected magnetic fields whereas, on field lines that reconnected earlier, the highest-speed flows have already passed by the observation location. This time asymmetry was clear in the E06 data[9].

To establish that interchange reconnection is the source of the bursty flows, we use the measurements and established principles of reconnection to deduce the basic characteristics in the low corona. The strength of the reconnecting magnetic field is a key parameter. As the field strength at the base of the corona has substantial variation, we estimate the amplitude of the reconnecting magnetic field by projecting the measured magnetic field at the PSP back to the solar surface. The R^{-2} fall-off of the radial magnetic field with heliospheric distance R is valid in the solar wind, but fails closer to the Sun. Thus, we use a combination of the R^{-2} behaviour at large R with a fall-off derived from a surface-averaged PFSS model below 2.5 R_s (Extended Data Fig. 1). The resulting projection of the 600-nT magnetic field at 13.4 R_s to the low corona is 4.5 G, which is consistent with the PFSS data in Fig. 2. Plasma density at the base of the corona is not measured directly. However, the characteristic amplitude of the bursty flows at PSP are around 300 km s^{-1}. Because the flows during bursty reconnection are Alfvénic, we can estimate the density knowing the magnetic field strength. The resulting density is around 10^9 cm^{-3}, a reasonable value for the low corona[31].

To address whether the rate of energy release is sufficient to drive, the wind we estimate the reconnection inflow rate V_r. A lower limit follows from the fact that the flow bursts are nearly continuous. We define the reconnection time as $t_r = L_B/V_r$, the time required for open field lines to traverse the characteristic scale length L_B of the surface magnetic field, which is around

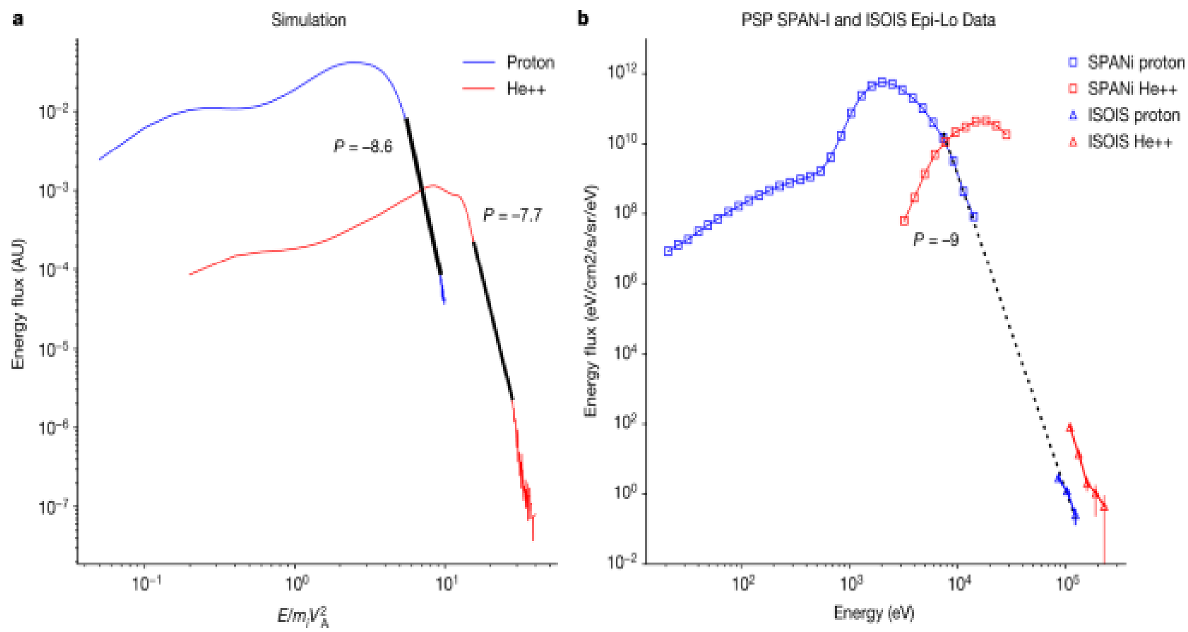
10° or 6×10^4 km. A second time is the time $t_b \approx R_{PSP}/V_R$ for the reconnection bursts to reach the spacecraft at R_{PSP}. In the limit $t_r >> t_b$, the outflows from the reconnection site would quickly pass by the spacecraft and there would be no high-speed flows until the spacecraft connected to another reconnection site. When $t_r \leq t_b$, the spacecraft would measure bursty flows as the spacecraft crossed the entire supergranulation scale. The observations show the latter because bursty flows are measured during the entire crossing of the supergranulation scale. Observations suggest that $t_r \leq t_b$ or $V_r \approx L_B V_R/R_{PSP}$ is ≈3 km s^{-1} or around 0.01 of the local Alfvén speed, a low value if reconnection is collisionless,[32,33,34] but comparable to the magnetohydrodynamic (MHD) prediction[35]. With ambient temperatures of around 100 eV, the reconnection electric field is therefore around three orders of magnitude above the Dreicer runaway field. In this regime, collisions are too weak to limit electron acceleration, and collisionless processes dominate. The rate of magnetic energy release from interchange reconnection is given by $V_r B^2/4\pi \approx 5 \times 10^5$ ergs cm^{-2} s^{-1} using $B = 4.5$ G and $V_r = 3$ km s^{-1}. This is comparable to that required to drive the high-speed wind, which is around 10^5–10^6 ergs cm^{-2} s^{-1}.

Thus through the PSP observations, the Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI) surface magnetic field measurements and well-known characteristics of magnetic reconnection, we have established that interchange reconnection is sufficient to drive both the ambient base solar wind flow through the radial pressure drop and the microstream bursts that lie on top of this flow. Further tests of the reconnection scenario concern the structuring of the flow bursts and the production of energetic protons and alphas. A key observation reported in the E06 data[9] and illustrated in the schematic in Fig. 3c is the temporal asymmetry in the bursts: large-amplitude bursts onset sharply and decrease gradually across the burst period and the time sequence then repeats. Data from a particle-in-cell (PIC) simulation is presented in Fig. 3b (Methods). A cut across the simulation outflow exhaust reveals high-speed bursts on newly reconnected field lines in the exhaust adjacent to the magnetic separatrix, whereas, on field lines in the exhaust interior, the fastest flow bursts have already passed the location of the cut, so the measured flows are weaker (Fig. 3a). Our simulations support the hypothesis that the bursts

observed by the PSP correspond to crossings of interchange reconnection exhausts. Dispersion signatures are well documented in the cusp of the Earth's magnetosphere as a result of reconnection at the terrestrial magnetopause[36]. Reconnection between the closed magnetic flux of the Earth and 'open' flux in the solar wind is an analogue of coronal interchange reconnection.

Finally, the spectrum of energetic protons and alphas has been calculated from the interchange reconnection simulations. The simulation includes fully stripped alpha particles that are 5% by number, similar to the solar atmosphere[16]. Energy flux spectra of both species are shown in Fig. 4a. Data is taken from the outflow exhaust and includes only plasma that has undergone acceleration. Protons and alphas exhibit an energetic, non-thermal power-law distribution with spectral indices of around −8 for both species. As shown in Fig. 4b, in the spectrum of the differential energy flux of particles during the time 04:00–19:00 on 20 November 2021 (from Fig. 1), there are also energetic protons and alphas with energies beyond 100 keV. The spectra are again rather soft, having spectral indices of around −9, consistent with the simulation data. The energy in the simulation is normalized to the free parameter $m_i {V}_{{\rm{A}}}^{2}$. By equating the energy minimum of the proton power law in the simulation (approximately 5 $m_i {V}_{{\rm{A}}}^{2}$) to that of the PSP measurements (approximately 7 keV), we find that the coronal value of $m_i {V}_{{\rm{A}}}^{2}$ is around 1.4 keV, compared with around 0.9 keV from the 300 km s^{-1} estimate for V_A based on the amplitude of the bursty flows measured at 13.4 R_s. That the two values of $m_i {V}_{{\rm{A}}}^{2}$ are close indicates that the Alfvén speed in the corona where reconnection is taking place is in the range of 300–400 km s^{-1}.

Fig. 4: Proton and alpha particle differential energy flux from interchange reconnection simulations and solar wind observations.



a, The proton (blue) and alpha particle (red) energy fluxes taken from the outflow exhaust from an interchange reconnection simulation (see Methods for simulation details). The energy normalization in the simulation is $m_i{V}_{{\rm{A}}}^{2}$, which is an arbitrary parameter[35]. The units in the ordinate direction are arbitrary, although the reduced height of the alpha flux reflects the 5% number density of alphas. Both fluxes peak and then roll over into distinct soft power laws with slopes of −8.6 and −7.7 for the protons and alphas, respectively, with the alpha spectrum shifted to higher energy than that of the protons. The spectral indices of the energetic ions depend on the magnitude of the ambient guide (out-of-plane) magnetic field with stronger guide fields producing softer spectra. The data are from a simulation with a guide field of 0.55 of the reconnecting magnetic field. **b**, The proton (blue) and alpha (red) energy fluxes from PSP measurements during the time interval 04:00:00–19:00:00 on 20 November 2021 from Fig. 1. As in the simulations, the spectra peak and roll over into power-law-like suprathermal tails with a similar slope of −9 for the protons. Thus, the power-law slopes from the simulation and the observational data are very close. The alpha measurement does not extend to high enough energy to characterize any power-law behaviour. Finally, we can use the low energy bound of the power-law distribution from the simulation (approximately 5 $m_i{V}_{{\rm{A}}}^{2}$) and observations (approximately 7 keV) to establish that the value of $m_i{V}_{{\rm{A}}}^{2}$ at the coronal

reconnection site is around 1.4 keV. This is comparable with around 0.9 keV from the 300 km s^{-1} estimate for V_A based on the amplitude of the bursty flows measured by the PSP at 13.4 R_s.

The picture that emerges is that reconnection directly heats the ambient coronal plasma sufficiently to drive the bulk outflow[37,38,39] and at the same time produces the turbulent velocity bursts that ride this outflow[24,25,26,27]. Extended Data Fig. 2 shows the strong heating of protons from the simulation in Fig. 3a,b. Of course, a fraction of the magnetic energy released during reconnection can take the form of Alfvén waves[1,2] or other magnetic structures[5] that can be dissipated higher in the coronal to further drive the bulk outflow[1,2,5,37]. However, the in situ data from reconnection in the Earth's magnetosphere[40] and at the heliospheric current sheet[41] show strong local plasma energization rather than wave generation. The time asymmetry that characterizes the bursty flows[9] and the spectral indices of the power-law distributions of energetic ions are in remarkably good agreement with the interchange reconnection simulation data, in which local plasma energization dominates waves and turbulence. Still, three-dimensional simulations with greater scale separation might show stronger magnetic turbulence. In either scenario, interchange reconnection is the likely energy drive mechanism of the fast solar wind. Recent remote-sensing measurements[42,43] also support the interchange magnetic reconnection scenario. We note that structured microstreams and magnetic switchbacks are present throughout the inner heliosphere measured by the PSP and that the primary difference between the slow and fast solar wind may lie in the magnetic topology of the underlying coronal hole.

Methods

Potential field source surface modelling

To generate the footpoints shown in Figs. 1e and 2, a PFSS model[16,17] was run using an Air Force Data Assimilative Photospheric Flux Transport–Global Oscillation Network Group (ADAPT–GONG) magnetogram[44,45,46] from 21 November 2022, with a source surface height set to the canonical

value of 2.5 R_S[19] by means of the open-source pfsspy[47] software. The footpoint mapping from the PSP down to the solar surface followed the methodology[48] comprising a ballistic heliosphere[49,50] and the PFSS domain from 2.5 R_S down to the photosphere[51].

The results for PSP E10 were distinct and compelling. As shown in Figs. 1 and 2, for 20–21 November, the PSP was rotating faster than the Sun and moving from left to right in the Carrington frame of reference shown in those plots. The footpoint mapping connected deep inside two mid-latitude negative polarity coronal holes of substantial area. This source mapping is uniquely well supported, compared with previous PSP encounters, owing to the comparison of the in situ data. First, the magnetic polarity measured by the PSP throughout the encounter is well explained by the PFSS current sheet geometry and coronal hole polarity. Second, the times when the PSP maps to the centre of these large coronal holes correspond to maxima in solar wind speed, and at the time the connection switches from one source to another in the model, there is a distinct dip in solar wind speed, clearly consistent with the traversal of overexpanded field lines at the coronal hole boundaries[52]. This correspondence is clearly shown in Fig. 1 where the transition between 'stream 1' and 'stream 2' marked in the bottom panel corresponds to the dip in solar wind speed (black trace, Fig. 1c).

PSP/SWEAP data analysis

We use proton and alpha particle measurements from the Solar Wind Electrons Alphas and Protons (SWEAP) instrument suite[11] on PSP. The proton spectrum in Fig. 4 is taken from the Solar Probe Analyzer (SPAN-Ion) SF00 data product, averaged over the time range from 20 November 2021 at 04:00:00 to 20 November 2021 at 19:00:00 and summed over all look directions. We work in units of energy flux as opposed to number flux or distribution function as it results in a spectrum spanning fewer orders of magnitude at high energy, facilitating the comparison between SPAN-Ion and Integrated Science Investigation of the Sun (ISOIS)/Epi-Lo data, as well as being the quantity most directly related to the SPAN-Ion measurements. The power law for the protons is fit to the four highest energy SPAN data points and the ISOIS data points. The alpha spectrum is obtained in the same way from SPAN-Ion's sf01 data product, except that a small amount (around

1%) of contaminant protons leaking in from the sf00 channel are accounted for and subtracted. The large shift to higher energy of the alphas relative to the protons during this interval means that the contaminant protons have no impact on the power law part of the spectrum or its exponent, and only affect the lowest energy data points.

Estimation of the reconnection magnetic field

The strength of the magnetic field that drives interchange reconnection will control the rate of magnetic energy release and the spectra of energetic particles produced. Although the SDO/HMI observations make known the structure of the magnetic field in the low corona, these measurements do not show the strength of the magnetic field that is actually undergoing reconnection because there is substantial variation of the field strength along the surface. To estimate the strength of the magnetic field driving the flow bursts measured by the PSP, we project the measured magnetic field at the PSP and project this magnetic field down to the solar surface. The radial magnetic field B_R at the perihelion of E10, as shown in Fig. 1, is around 600 nT. Direct measurements of the radial profile of B_R over the first five PSP orbits have established an R^{-2} scaling for the field[45], consistent with the conservation of the radial magnetic flux. However, deviations from this scaling are expected close to the Sun. Specifically, because closed flux occupies a substantial fraction of the solar surface, the open flux will be compressed into a reduced fraction of the solar surface, which will lead to greater compression of the magnetic field near the solar surface. A rough estimate of the increased magnetic field compression can be obtained by averaging the radial magnetic field obtained from the PFSS model during the E10 perihelion. The radial dependence of this averaged field is shown in Extended Data Fig. 1. The magnetic field compression from 2.5 R_s, the outer boundary of the PFSS grid, down to just above the solar surface is around 26, which is well above the compression of around 6.25 from the R^{-2} dependence. Thus, we assume that the R^{-2} describes the radial dependence from $R = 13.4\ R_s$ to 2.5 R_s and take the compression of 26 from 2.5 R_s to just above the solar surface. The projection of the 600-nT field down to the solar surface is around 4.5 G, which is in reasonable agreement with the strength of the solar surface magnetic field shown in Fig. 2.

Particle-in-cell simulations

Our estimate of the rate of interchange reconnection based on projections of the PSP observations back to the low corona suggest that reconnection there is deeply in the collisionless regime. To explore the structure of the interchange reconnection exhaust and the resulting energetic proton and alpha spectra measured at the PSP, we use the PIC model p3d (ref. [53]). The MHD model is not adequate to explore the particle energization documented in the PSP data. We limit the calculations to a two-dimensional system with an initial magnetic geometry that leads to reconnection between open and closed flux low in the corona[3,54]. Because of constraints on the domain size possible with the PIC model, there is no gravity in the simulations, so the model does not describe the complete dynamics of the solar wind drive mechanism. In addition, line-tied boundary conditions are not imposed at the nominal coronal surface. Thus, the model is not a complete description of interchange reconnection in the low corona, but will provide information on the dynamics of collisionless reconnection, the structure of the outflow exhaust, the bulk heating and the spectra of accelerated particles. We include alpha particles (5% by number) so that the spectra of protons and alphas can be compared.

The initial state for the simulation consists of a band of vertical flux (field strength B_0 in the negative radial direction) with a low plasma density (0.1 n_0) and an adjacent region with higher density that is a cylindrical equilibrium. The detailed initial state has been described previously[23], so the governing equations are not repeated here. The peak magnetic field of the cylindrical equilibrium is 0.76 B_0 with a peak density of n_0. The temperatures are uniform with $T_e = T_p = T_a = 0.06\, m_p {V}_{{\rm{A}}0}^{2}$ with V_{A0} the Alfvén speed based on B_0, n_0 and the proton mass m_p. Thus, in the initial state, the plasma pressure is small compared with the magnetic pressure, as expected in the corona. The guide field B_z is non-zero everywhere with a profile that balances the pressure and tension forces. The strength of the guide field can be varied by choosing its value in the region of vertical flux. Its value does not substantially affect the overall structure and dynamics of reconnection shown in Fig. 3. However, because a strong guide field weakens the Fermi drive mechanism for particle

energy gain, the guide field controls the power-law index of energetic protons and alphas. The energy fluxes shown in Fig. 4 were from a simulation with a guide field of 0.55 B_0. Simulations with a weaker (stronger) guide field produced harder (softer) spectra.

The results of the simulation are presented in normalized units: times to the Alfvén transit time across the domain of scale length L, L/V_{A0}, plasma fluxes to n_0V_{A0} and energies to $m_p{V}_{{\rm{A}}0}^{2}$. The domain dimensions in the x and y directions are equal. The mass ratio $m_p/m_e = 25$ is artificial as is the velocity of light (20 V_{A0}) and the proton inertial scale $d_p = L/163.84$. As has been established in earlier papers, the results are not sensitive to these values[33,34]. The radius of the cylindrical magnetic field is 60 d_p and the grid scales are 0.02 d_p in both space directions, with around 400 particles per cell.

Because the velocities and energies in the simulation are normalized to the Alfvén speeds V_{A0} and $m_p{V}_{{\rm{A}}0}^{2}$, respectively, direct comparison with the observations requires that these parameters, and specifically the Alfvén speed V_{A0} where reconnection is taking place, be established. As described in the main text, we use two distinct approaches for estimating V_{A0}. The first comes from the amplitude of the flow bursts measured by the PSP at 13.4 R_s, which have values of around 300 km s^{-1}. The second comes from comparing the proton spectra from the simulation and that measured with SPAN-Ion. Specifically, we equate the low energy limit of the proton power-law spectrum from the simulation and the observations. This yields $m_p{V}_{{\rm{A}}0}^{2} = 1.4$ keV, which corresponds to $V_{A0} = 370$ km s^{-1}. Thus, the two approaches yield comparable values, which enables us to directly compare the simulation results with the observations. The spectral indices of around −8 for the proton and alpha energy fluxes from the simulation are independent of this normalization. That they are in approximate agreement with the observational data is strong support for the interchange reconnection model for these energetic particles. We further note that a simulation with half of the domain size ($L = 81.92\ d_p$) produced power-law spectra with similar spectral indices.

Many of the global models that have been used to explore the acceleration of the solar wind have been based on the assumption that Alfvén waves[1,2] or other forms of magnetic structures[5] are injected into the low corona and the heating associated with this turbulence produces the pressure required to drive the wind. As interchange reconnection is often invoked as the source of this turbulence, it is important to explore whether a substantial fraction of released magnetic energy appears as magnetic turbulence versus direct particle energization or bulk flow. Shown in Extended Data Fig. 2 is a two-dimensional plot of the proton temperature from the same simulation and at the same time as the plot of the vertical proton flux in Fig. 3b. The entire outflow exhaust is filled with high temperature protons with temperatures that are a large fraction of $m_{\rm p}{V}_{{\rm{A}}0}^{2}$, which, as discussed in the main text, is in the range of 0.9 keV to 1.4 keV. The warped field lines in the reconnection exhaust that are evident in Fig. 3b and Extended Data Fig. 2 suggest that reconnection also drives waves and magnetic turbulence. A detailed exploration of the relative energy that appears as magnetic turbulence versus direct heating has not been carried out but is an important extension of the present results, especially in three dimensions, where the process of reconnection is much more dynamic.

Data availability

The PSP mission data used in this study are openly available at the NASA Space Physics Data Facility (https://nssdc.gsfc.nasa.gov) and were analysed using the IDL/SPEDAS software package (https://spedas.org/blog/). The computer simulations used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract no. DE-AC02- 05CH11231. Simulation data is available at https://doi.org/10.5281/zenodo.7562035.

References

1. McKenzie, J. F., Banaszkiewicz, M. & Axford, W. I. Acceleration of the high speed solar wind. *Astron. Astrophys.* **303**, L45 (1995).

2. Axford, W. I. et al. Acceleration of the high speed solar wind in coronal holes. *Space Sci. Rev.* **98**, 25 (1999).

3. Fisk, L. A., Schwadron, N. A. & Zurbuchen, T. H. Acceleration of the fast solar wind by the emergence of new magnetic flux. *J. Geophys. Res.* **104**, 19765 (1999).

4. Cranmer, S. R. & van Ballegooijen, A. A. Can the solar wind be driven by magnetic reconnection in the Sun's magnetic carpet? *Astrophys. J.* **720**, 824 (2010).

5. Zank, G. P. et al. Theory and transport of nearly incompressible magnetohydrodynamic turbulence. IV. Solar coronal turbulence. *Astrophys. J.* **854**, 32 (2018).

6. Fox, N. J. et al. The Solar Probe Plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7 (2016).

7. Bale, S. D. et al. Highly structure slow solar wind emerging from an equatorial coronal hole. *Nature* **576**, 237 (2019).

8. Kasper, J. C. et al. Alfvénic velocity spikes and rotational flows in the near-Sun solar wind. *Nature* **576**, 228 (2019).

9. Bale, S. D. et al. A solar source of Alfvénic magnetic field switchbacks: in situ remnants of magnetic funnels on supergranulation scales. *Astrophys. J.* **923**, 174 (2021).

10. Rieutord, M. & Rincon, F. The Sun's supergranulation. *Living Rev. Sol. Phys.* **7**, 2 (2010).

11. Kasper, J. C. et al. Solar Wind Electrons Alphas and Protons (SWEAP) investigation: design of the solar wind and coronal plasma instrument suite for Solar Probe Plus. *Space Sci. Rev.* **204**, 131 (2016).

12. McComas, D. J. et al. Integrated Science Investigation of the Sun (ISIS): design of the energetic particle investigation. *Space Sci. Rev.* **204**, 187 (2016).

13. Bale, S. D. et al. The FIELDS instrument suite for Solar Probe Plus: measuring the coronal plasma and magnetic field, plasma waves and turbulence, and radio signatures of solar transients. *Space Sci. Rev.* **204**, 49 (2016).

14. Thieme, K. M., Marsch, E. & Schwenn, R. Spatial structures in high-speed streams as signatures of fine structures in coronal holes. *Ann. Geophys.* **8**, 713 (1990).

15. Neugebauer, M. et al. Ulysses observations of microstreams in the solar wind from coronal holes. *J. Geophys. Res.* **100**, 23389 (1995).

16. Kasper, J. C. et al. Solar wind helium abundance as a function of speed and heliographic latitude: variation through a solar cycle. *Astrophys. J.* **660**, 901 (2007).

17. Schatten, K. H., Wilcox, J. M. & Ness, N. F. A model of interplanetary and coronal magnetic fields. *Solar Phys.* **6**, 442 (1969).

18. Altschuler, M. D. & Newkirk, G. Magnetic fields and the structure of the solar corona. 1: methods of calculating coronal fields. *Solar Phys.* **9**, 131 (1969).

19. Hoeksema, J. T. *Structure and Evolution of the Large Scale Solar and Heliospheric Magnetic Fields*. PhD thesis, Stanford Univ. (1984).

20. Lemen, J. R. et al. The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 17 (2012).

21. Fisk, L. A. & Kasper, J. C. Global circulation of the open magnetic flux of the Sun. *Astrophys. J. Lett.* **894**, L4 (2020).

22. Zank, G. P. et al. The origin of switchbacks in the solar corona: Linear theory. *Astrophys. J.* **903**, 1 (2020).

23. Drake, J. F. et al. Switchbacks as signatures of magnetic flux ropes generated by interchange reconnection in the corona. *Astron. Astrophys.* **650**, A2 (2021).

24. Drake, J. F. et al. Formation of secondary islands during magnetic reconnection. *Geophys. Res. Lett.* **33**, 13015 (2006).

25. Bhattacharjee, A. et al. Fast reconnection in high-Lundquist-number plasmas due to the plasmoid instability. *Phys. Plasmas* **16**, 112102 (2009).

26. Cassak, P. A., Shay, M. A. & Drake, J. F. Scaling of Sweet–Parker reconnection with secondary islands. *Phys. Plasmas* **16**, 120702 (2009).

27. Daughton, W. et al. Transition from collisional to kinetic regimes in large-scale reconnection layers. *Phys. Rev. Lett.* **103**, 065004 (2009).

28. Drake, J. F. et al. Ion heating resulting from pickup in magnetic reconnection exhausts. *J. Geophys. Res.* **114**, A05111 (2009).

29. Drake, J. F., Swisdak, M. & Fermo, R. The power-law spectra of energetic particles during multi-island magnetic reconnection. *Astrophys. J. Lett.* **763**, L5 (2013).

30. Zhang, Q., Guao, F. & Daughton, W. Efficient nonthermal ion and electron acceleration enabled by the flux-rope kink instability in 3D nonrelativistic magnetic reconnection. *Phys. Rev. Lett.* **127**, 185101 (2021).

31. Cranmer, S. R. et al. An empirical model of a polar coronal hole at solar minimum. *Astrophys. J.* **511**, 481 (1999).

32. Shay, M. A. et al. The scaling of collisionless magnetic reconnection for large systems. *Geophys. Res. Lett.* **26**, 2163 (1999).

33. Shay, M. A., Drake, J. F. & Swisdak, M. Two-scale structure of the electron dissipation region during collisionless magnetic reconnection. *Phys. Rev. Lett.* **99**, 155022 (2007).

34. Daughton, W. et al. Computing the reconnection rate in turbulent kinetic layers by using electron mixing to identify topology. *Phys. Plasmas* **21**, 052307 (2014).

35. Huang, Y.-M. & Bhattacharjee, A. Turbulent magnetohydrodynamic reconnection mediated by the plasmoid instability. *Astrophys. J.* **818**, 20 (2016).

36. Burch, J. L. et al. Plasma injection and transport in the mid-latitude polar cusp. *Geophys. Res. Lett.* **9**, 921 (1982).

37. Withbroe, G. L. The temperature structure, mass, and energy flow in the corona and inner solar wind. *Astrophys. J.* **325**, 442 (1988).

38. Parker, E. N. Dynamics of interplanetary gas and magnetic fields. *Astrophys. J.* **128**, 664 (1958).

39. Axford, W. I. & McKenzie, J. F. The origin of high speed solar wind streams. In *Solar Wind Seven*(eds Marsch, E. & Schwenn, R.) 1–5 (Pergamon Press, 1992).

40. Ergun, R. E. et al. Observations of particle acceleration in magnetic reconnection-driven turbulence. *Astrophys. J.* **898**, 154 (2020).

41. Phan, T. D. et al. Parker Solar Probe observations of solar wind energetic proton beams produced by magnetic reconnection in the near-Sun heliospheric current sheet. *Geophys. Res. Lett.* **49**, e96986 (2022).

42. Telloni, D. et al. Observation of a magnetic switchback in the solar corona. *Astrophys. J. Lett.* **936**, L25 (2022).

43. Raouafi, N. E. et al. Magnetic reconnection as the driver of the solar wind. *Astrophys. J.***945**, 28 (2023).

44. Arge, C. N. et al. Air Force Data Assimilative Photospheric Flux Transport (ADAPT) model. In *AIP Confererence Proceedings* (eds Maksimovic, M. et al) **1216**, 343 (AIP, 2010).

45. Arge, C. N. et al. Improving data drivers for coronal and solar wind models. In *5th International Conference of Numerical Modeling of Space Plasma Flows (ASTRONUM 2010)* (eds Pogorelov, N. V. et al) **444**, 99–104 (ASP, 2011).

46. Arge, C. N. et al. Modeling the corona and solar wind using ADAPT maps that include far-side observations. In *AIP Conference Proceedings* **1539**, 11–14 (AIP Publishing LLC, 2013).

47. Stansby, D., Yeates, A. & Badman, S. pfsspy: a Python package for potential source surface modeling. *J. Open Source Software* **5**, 2732 (2020).

48. Badman, S. T. et al. Magnetic connectivity of the ecliptic plane within 0.5 au: potential field source surface modeling of the first Parker Solar Probe encounter. *Astrophys. J. Supp.* **246**, 23 (2020).

49. Nolte, J. T. & Roelof, E. C. Large-scale structure of the interplanetary medium, I: high coronal source longitude of the quiet-time solar wind. *Solar Phys.* **33**, 241 (1973).

50. MacNeil, A. R. et al. A statistical evaluation of ballistic backmapping for the slow solar wind: the interplay of solar wind acceleration and corotation. *MNRAS* **10**, 1093 (2021).

51. Badman, S. T. et al. Prediction and verification of Parker Solar Probe solar wind sources at 13.3 R_S. *J. Geophys. Res.* **128**, e2023JA031359 (2023).

52. Wang, Y.-M. & Sheeley, N. R. Jr Solar wind speed and coronal flux-tube expansion. *Astrophys. J.* **355**, 726 (1990).

53. Zeiler, A. et al. Three-dimensional particle simulations of collisionless magnetic reconnection. *J. Geophys. Res.* **107**, 1230 (2002).

54. Fisk, L. A. The open magnetic flux of the Sun. I. Transport by reconnections with coronal loops. *Astrophys. J.* **636**, 563 (2005).

Acknowledgements

The FIELDS, SWEAP and ISOIS suites were designed, developed and are operated under NASA contract NNN06AA01C. We acknowledge the extraordinary contributions of the PSP mission operations and spacecraft

engineering team at the Johns Hopkins University Applied Physics Laboratory. M.V. was supported in part by the International Space Science Institute, Bern, through the J. Geiss fellowship. J.F.D. and M.S. were supported by the NASA Drive Science Center on Solar Flare Energy Release (SolFER) under Grant 80NSSC20K0627, NASA Grant 80NSSC22K0433 and NSF Grant PHY2109083. T.S.H. is supported by STFC grant ST/W001071/1. O.P. was supported by the NASA Grant 80NSSC20K1829. Elements of this work benefited from discussions at the meeting of Team 463 at the International Space Science Institute (ISSI).

Author information

Authors and Affiliations

1. Physics Department, University of California, Berkeley, CA, USA

   S. D. Bale & M. D. McManus

2. Space Sciences Laboratory, University of California, Berkeley, CA, USA

   S. D. Bale, M. D. McManus, D. E. Larson & T. Phan

3. Department of Physics, the Institute for Physical Science and Technology and the Joint Space Institute, University of Maryland, College Park, MD, USA

   J. F. Drake

4. Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD, USA

   J. F. Drake & M. Swisdak

5. Southwest Research Institute, San Antonio, TX, USA

   M. I. Desai

6. Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

S. T. Badman

7. The Blackett Laboratory, Imperial College London, London, UK

T. S. Horbury

8. Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA

N. E. Raouafi & D. Mitchell

9. Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA, USA

M. Velli

10. International Space Science Institute, Bern, Switzerland

M. Velli

11. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

D. J. McComas

12. California Institute of Technology, Pasadena, CA, USA

C. M. S. Cohen

13. Advanced Heliophysics Inc., Los Angeles, CA, USA

O. Panasenco

14. BWX Technologies, Inc., Washington, DC, USA

J. C. Kasper

15. Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

J. C. Kasper

Contributions

S.D.B. and J.F.D. wrote the manuscript with major contributions from S.T.B. S.D.B. analysed the PSP measurements, with contributions from M.D.M., M.I.D., T.S.H. and D.E.L. S.T.B. performed the PFSS analysis. J.F.D. and M.S. performed the computer simulations. S.D.B., J.C.K. and D.J.M. lead the PSP/FIELDS, SWEAP and ISOIS teams, respectively. All authors participated in the data interpretation and read and commented on the manuscript.

Corresponding author

Correspondence to S. D. Bale.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Vadim Uritsky, G. P. Zank and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Magnitude of the radial magnetic field from PFSS.

The magnitude of the radial component of the magnetic field is modelled by the PFSS implementation, constrained by magnetograms at the photospheric footpoints and the open boundary condition at 2.5 R_S. This field magnitude is consistent with the field measured at Parker Solar Probe and is used to estimate the Alfvén speed at the reconnection site.

Extended Data Fig. 2 Proton temperature in the plane of reconnection with overlying magnetic field lines.

The proton temperature is shown from the same simulation and the same time as the vertical flux shown in Fig. 3b. The temperature is normalized to $m_p{V}_{{\rm{A}}{\rm{0}}}^{{\rm{2}}}$, which, as discussed in the main text is in the range of 0.9 keV to 1.6 keV. Thus, since the magnetic energy released per particle during reconnection is around $m_p{V}_{{\rm{A}}{\rm{0}}}^{{\rm{2}}}$, a large fraction of the released magnetic energy is goes into the heating and energization of the ambient plasma very close to the magnetic energy release site. The energy going into turbulent magnetic fields that is injected upward into the corona and is available to further heat the ambient plasma remains to be explored.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from

the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Bale, S.D., Drake, J.F., McManus, M.D. *et al.* Interchange reconnection as the source of the fast solar wind within coronal holes. *Nature* **618**, 252–256 (2023). https://doi.org/10.1038/s41586-023-05955-3

- Received: 11 August 2022
- Accepted: 14 March 2023
- Published: 07 June 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05955-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

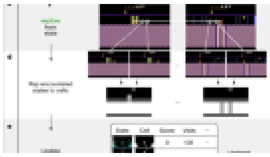
A brain circuit for infanticide, in mice

- Noah Baker

- Nick Petrić Howe

Nature Nature Podcast 07 Jun 2023

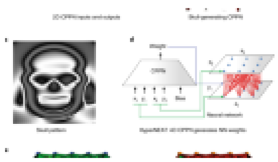
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05955-3



First return, then explore

24 February 2021

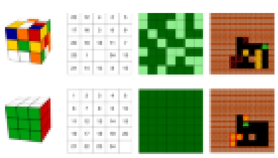
Adrien Ecoffet, Joost Huizinga, … Jeff Clune



Designing neural networks through neuroevolution

07 January 2019

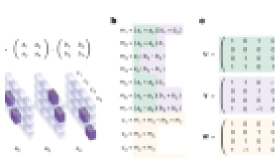
Kenneth O. Stanley, Jeff Clune, … Risto Miikkulainen



Solving the Rubik’s cube with deep reinforcement learning and search

15 July 2019

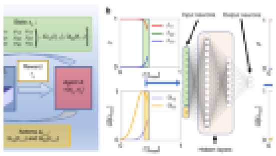
Forest Agostinelli, Stephen McAleer, … Pierre Baldi



[Discovering faster matrix multiplication algorithms with reinforcement learning]

05 October 2022

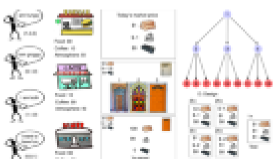
Alhussein Fawzi, Matej Balog, … Pushmeet Kohli



[Coherent transport of quantum states by deep reinforcement learning]

20 June 2019

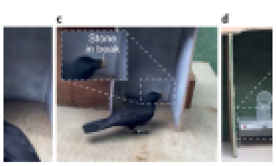
Riccardo Porotti, Dario Tamascelli, … Enrico Prati



[Multi-task reinforcement learning in humans]

28 January 2021

Momchil S. Tomov, Eric Schulz & Samuel J. Gershman



[Intelligent problem-solving as integrated hierarchical reinforcement learning]

25 January 2022

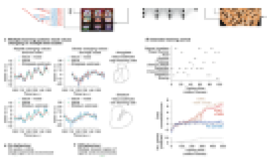
Manfred Eppe, Christian Gumbsch, … Stefan Wermter



Evolutionary training and abstraction yields algorithmic generalization of neural computers

16 November 2020

Daniel Tanneberg, Elmar Rueckert & Jan Peters



Reinforcement learning in artificial and biological systems

04 March 2019

Emre O. Neftci & Bruno B. Averbeck

- Article
- Open Access
- Published: 07 June 2023

Faster sorting algorithms discovered using deep reinforcement learning

- Daniel J. Mankowitz ,
- Andrea Michi,
- Anton Zhernov,

- Marco Gelmi,
- Marco Selvi,
- Cosmin Paduraru,
- Edouard Leurent,
- Shariq Iqbal,
- Jean-Baptiste Lespiau,
- Alex Ahern,
- Thomas Köppe,
- Kevin Millikin,
- Stephen Gaffney,
- Sophie Elster,
- Jackson Broshear,
- Chris Gamble,
- Kieran Milan,
- Robert Tung,
- Minjae Hwang,
- Taylan Cemgil,
- Mohammadamin Barekatain,
- Yujia Li,
- Amol Mandhane,
- Thomas Hubert,
- Julian Schrittwieser,
- Demis Hassabis,
- Pushmeet Kohli,
- Martin Riedmiller,
- Oriol Vinyals &
- …
- David Silver

Show authors

Nature **volume 618**, pages 257–263 (2023)

- 201k Accesses
- 972 Altmetric
- Metrics details

Abstract

Fundamental algorithms such as sorting or hashing are used trillions of times on any given day[1]. As demand for computation grows, it has become critical for these algorithms to be as performant as possible. Whereas remarkable progress has been achieved in the past[2], making further improvements on the efficiency of these routines has proved challenging for both human scientists and computational approaches. Here we show how artificial intelligence can go beyond the current state of the art by discovering hitherto unknown routines. To realize this, we formulated the task of finding a better sorting routine as a single-player game. We then trained a new deep reinforcement learning agent, AlphaDev, to play this game. AlphaDev discovered small sorting algorithms from scratch that outperformed previously known human benchmarks. These algorithms have been integrated into the LLVM standard C++ sort library[3]. This change to this part of the sort library represents the replacement of a component with an algorithm that has been automatically discovered using reinforcement learning. We also present results in extra domains, showcasing the generality of the approach.

Main

Human intuition and know-how have been crucial in improving algorithms. However, many algorithms have reached a stage whereby human experts have not been able to optimize them further, leading to an ever-growing computational bottleneck. The work in classical program synthesis literature, spanning many decades, aims to generate correct programs and/or optimize programs using proxies for latency. These include enumerative search techniques[4,5,6,7] and stochastic search[5,6,8,9,10] as well as the more recent trend of using deep learning in program synthesis for generating correct programs[11,12,13,14,15,16]. Using deep reinforcement learning (DRL), we can take this a step further by generating correct and performant algorithms by optimizing for actual measured latency at the CPU instruction level, by more efficiently searching and considering the space of correct and fast programs compared to previous work.

One of the fundamental questions in computer science is how to sort a sequence[17,18,19,20]. This is taught in elementary computer science classes around the world[21,22] and is used ubiquitously by a vast range of applications[23,24,25]. Decades of computer science research have focused on discovering and optimizing sorting algorithms[26,27,28]. A key component of practical solutions is a small sort over a short sequence of elements; this algorithm is called repeatedly when sorting large arrays that use divide-and-conquer approaches[29]. In this work, we focus on two types of small sort algorithm: (1) the fixed sort and (2) the variable sort. Fixed sort algorithms sort sequences of a fixed length (for example, sort 3 can only sort sequences of length 3), whereas variable sort algorithms can sort a sequence of varying size (for example, variable sort 5 can sort sequences ranging from one to five elements).

We formulate the problem of discovering new, efficient sorting algorithms as a single-player game that we refer to as AssemblyGame. In this game, the player selects a series of low-level CPU instructions, which we refer to as assembly instructions[30], to combine to yield a new and efficient sorting algorithm. This is challenging as the player needs to consider the combinatorial space of assembly instructions to yield an algorithm that is both provably correct and fast. The hardness of the AssemblyGame arises not only from the size of the search space, which is similar to extremely challenging games such as chess (10^{120} games)[31] and Go (10^{700} games)[32], but also from the nature of the reward function. A single incorrect instruction in the AssemblyGame can potentially invalidate the entire algorithm, making exploration in this space of games incredibly challenging.

To play the game, we introduce AlphaDev, a learning agent that is trained to search for correct and efficient algorithms. This agent is comprised of two core components, namely (1) a learning algorithm and (2) a representation function. The AlphaDev learning algorithm can incorporate both DRL as well as stochastic search optimization algorithms to play AssemblyGame. The primary learning algorithm in AlphaDev is an extension of AlphaZero[33], a well-known DRL algorithm, in which a neural network is trained to guide a search to solve AssemblyGame. The representation function is interchangeable and captures the underlying structure of

assembly programs. The primary AlphaDev representation is based on Transformers[34].

Using AlphaDev, we have discovered fixed and variable sort algorithms from scratch that are both new and more efficient than the state-of-the-art human benchmarks. The fixed sort solutions for sort 3, sort 4 and sort 5 discovered by AlphaDev have been integrated into the standard sort function in the LLVM standard C++ library[3]. This library is used by several million users including universities and numerous international companies[35]. In addition, we analyse the new algorithm discoveries, compare AlphaDev to stochastic search optimization approaches and apply AlphaDev to further domains to showcase the generality of the approach.

Representing algorithms as low-level CPU instructions

When compiling algorithms to machine code from a high level language such as C++ (for example, the sorting function in Fig. 1a), the algorithm is first compiled into assembly (Fig. 1b). The assembler then converts the assembly program into executable machine code. In this work, we optimize algorithms at the assembly level[30]. In a typical assembly program, the values are copied from memory into registers, manipulated between registers and then written back to memory. The set of assembly instructions supported depends on the processor architecture. For the purposes of this work, we focus on a subset of assembly instructions supported by the x86 processor architecture using the AT&T syntax[36]. Each instruction is of the format Opcode⟨Operand_A, Operand_B⟩. An example instruction is mov, which is defined as move a value from source (A) to destination (B). Further instruction definitions such as compare (cmp), conditional move (cmovX) and jump (jX) can be found in Extended Data Table 1. In the example in Fig. 1b, %eax, %ecx, %edx, %edi correspond to four different register locations and (%rsi), 4(%rsi) correspond to two different memory locations. The symbol $2 is a placeholder for a constant value, which corresponds to the length of the vector in this example. We use the terms assembly program and assembly algorithm interchangeably in this work. This is because AlphaDev builds an assembly program from scratch, from an initially unordered set of

instructions, each time it plays AssemblyGame, defining a new and efficient algorithm.

Fig. 1: The relationship between C++ and assembly programs.

a

```cpp
void variable_sort_2(int length, int *a)
{
  // Determine the number of elements
  switch (length)
  {
    case 0:
    case 1:
      // Exit if less than 2 elements
      return;
    case 2:
      // Below routine sorts 2 elements
      int tmp = a[0];
      a[0] = (a[1] < a[0]) ? a[1] : a[0];
      a[1] = (a[1] < tmp) ? tmp : a[1];
      return;
  }
}
```

b

```asm
variable_sort_2(int, int*):
        ; Determine the number of elements
        cmp    $2, %edi
        ; Exit if less than 2 elements
        jne     .Label
        ; Below routine sorts 2 elements
        mov    (%rsi), %eax
        mov    4(%rsi), %ecx
        cmp    %eax, %ecx
        mov    %eax, %edx
        cmovl   %ecx, %edx
        mov    %edx, (%rsi)
        cmovg %ecx, %eax
        mov    %eax, 4(%rsi)

.Label:
        retq
```

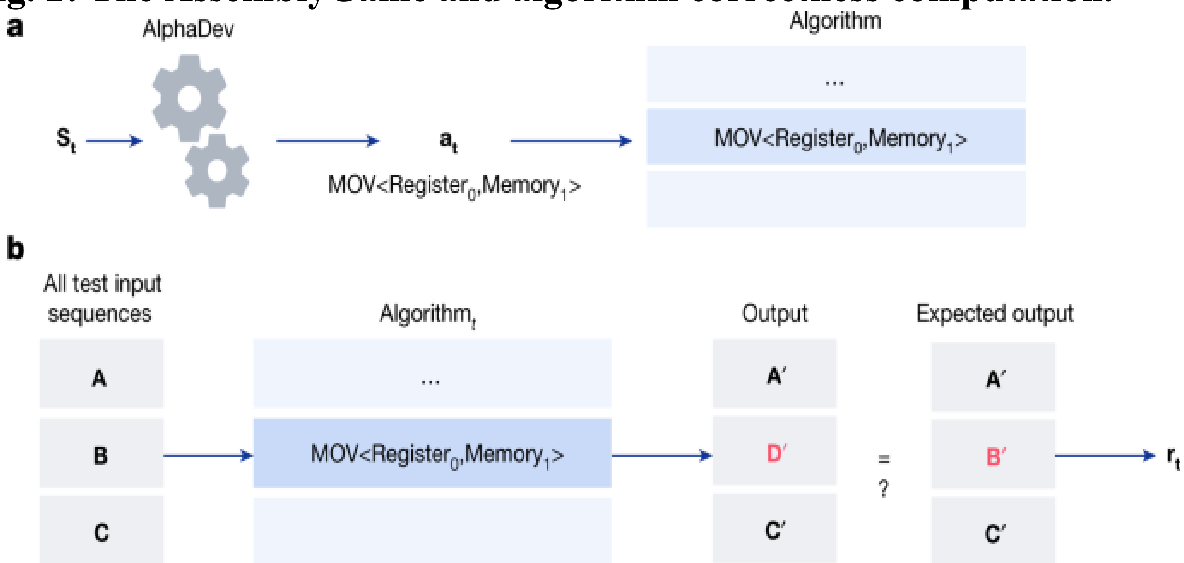
a, A C++ implementation of a variable sort 2 function that sorts any input sequence of up to two elements. **b**, The C++ implementation in **a** is compiled to this equivalent low-level assembly representation.

DRL for discovering faster algorithms

In this section, we formulate optimizing algorithms at the CPU instruction level as a reinforcement learning (RL) problem[37], in which the environment is modelled as a single-player game that we refer to as AssemblyGame. Each state in this game is defined as a vector $\mathbf{S_t} = \langle \mathbf{P_t}, \mathbf{Z_t} \rangle$ where $\mathbf{P_t}$ is a representation of the algorithm generated thus far in the game and $\mathbf{Z_t}$ represents the state of memory and registers after executing the current algorithm on a set of predefined inputs. As seen in Fig. 2a, at timestep t, the player receives the current state $\mathbf{S_t}$ and executes an action $\mathbf{a_t}$. This involves appending a legal assembly instruction (for example, mov<A,B>) to the current algorithm generated thus far. A reward $r_t$is received that comprises both a measure of algorithm correctness and latency. Algorithm correctness (Fig. 2b) involves inputting a set of N test sequences into the current algorithm $\mathbf{P_t}$ to generate N outputs. These outputs are then compared to the expected outputs and a correctness reward r_t is computed. Latency rewards

can be generated by either (1) penalizing the agent for increasing the length of the algorithm (when length and latency are highly correlated) that we refer to as the algorithm length reward, or (2) measuring the actual latency of the algorithm. The game is executed for a limited number of steps, after which the game is terminated. Winning the game corresponds to generating a correct, low-latency algorithm using assembly instructions. Losing the game corresponds to generating an incorrect algorithm or a correct but inefficient algorithm.

Fig. 2: The AssemblyGame and algorithm correctness computation.



a, The AssemblyGame is played by AlphaDev, which receives as input the current assembly algorithm generated thus far $\mathbf{S_t}$ and plays the game by selecting an action to execute. In this example, the action is a mov<Register$_0$,Memory$_1$> assembly instruction, which is appended to the current algorithm. The agent receives a reward that is a function of the algorithm's correctness, discussed in **b**, as well as the algorithm's latency. The game is won by the player discovering a low latency, correct algorithm. **b**, The program correctness and latency computations are used to compute the reward r_t. In this example, test sequences are input to the algorithm; for example, in the case of sorting three elements, test inputs comprise all sequences of unsorted elements of length 3. For each sequence, the algorithm output is compared to the expected output (in the case of sorting, the expected output is the sorted elements). In this example, the output \ ({\bf{D}}{\boldsymbol{{\prime} }}\) does not match the expected output \

({\bf{B}}{\boldsymbol{{\prime} }}\) and the algorithm is therefore incorrect.

We refer to the agent that plays this single-player game as AlphaDev. The agent's primary learning algorithm is an extension of the AlphaZero agent[32] and guides a Monte Carlo tree search (MCTS) planning procedure using a deep neural network[33,38]. The input to the neural network is the state $\mathbf{S_t}$ and the output is a policy and value prediction. The policy prediction is a distribution over actions and the value function is a prediction of the cumulative returns R that the agent should expect to receive from the current state $\mathbf{S_t}$. During a game, the agent receives as input the current state $\mathbf{S_t}$. The agent then executes an MCTS procedure and uses this to select the next action to take. The generated games are then used to update the network's parameters, enabling the agent to learn.

It is critical that AlphaDev has a representation[39,40] capable of representing complex algorithmic structures to efficiently explore the space of instructions. To achieve this, we introduce the AlphaDev representation network (Extended Data Fig. 1a). This network comprises two components, namely (1) a transformer encoder network that provides the agent with a representation of the algorithm structure, and (2) the CPU state encoder network that helps the agent predict how the algorithm affects the dynamics of memory and registers. The CPU state encoder network comprises a multilayer perceptron that receives as input the state of each register and memory location for a given set of inputs. These networks each output embeddings that are combined to yield the AlphaDev state representation.

Transformer encoder

Transformers are natural text encoders and have had much success with language models recently[14,34,41]. As such, this motivated us to adapt the standard transformer to model assembly instructions. We developed and incorporated a transformer encoder, our adaptation of the MultiQuery transformer encoder[42], into the AlphaDev representation network to represent the assembly instructions. Each assembly instruction's Opcode and corresponding Operands are converted to one-hot encodings and concatenated to form the raw input sequence. This is fed through a

multilayer transformer encoder, which maps it to corresponding embedding vectors (see Extended Data Fig. 1b for an illustration).

Latency value functions

Latency is an important reward signal that is used to guide the agent in discovering performant algorithms. To better estimate latency, we implemented a dual value function setup, whereby AlphaDev has two value function heads: one predicting algorithm correctness and the second predicting algorithm latency. The latency head is used to directly predict the latency of a given program by using the program's actual computed latency as a Monte Carlo target for AlphaDev during training. This dual-head approach achieved substantially better results than the vanilla, single head value function setup when optimizing for real latency.

Results

Discovering faster sort algorithms

We trained the AlphaDev agent from scratch to generate a range of fixed sort and variable sort algorithms that are both correct and achieve lower latency than the state-of-the-art human benchmarks.

Fixed sorting algorithms

We considered three fundamental algorithms: sort 3, sort 4 and sort 5. The state-of-the-art human benchmarks for these algorithms are sorting networks[43] as they generate efficient, conditional branchless assembly code. This means that all instructions are executed sequentially and there is no branching involved. Improving on these algorithms is challenging as they are already highly optimized. As seen in Table 1a, AlphaDev is able to find algorithms with fewer instructions than the human benchmarks for sort 3 and sort 5 and matches the state-of-the-art performance on sort 4. These shorter algorithms do indeed lead to lower latency as the algorithm length and latency are correlated for the conditional branchless case; see Appendix B in Supplementary Information for more details. We also explored scaling

to slightly larger sorts using a variant of AlphaDev. We managed to save three instructions on sort 6, two instructions on sort 7 and one instruction on sort 8, which provides a promising basis for future work. See Appendix C in Supplementary Information for an overview of the approach.

Table 1 AlphaDev performance when optimizing for algorithm length and latency

Variable sorting algorithms

We considered three variable sorting algorithms: VarSort3, VarSort4 and VarSort5. The human benchmark in each case is defined as an algorithm that, for a given input length, calls the corresponding sorting network. In this case, branching is required, which greatly increases the complexity of the problem as the agent needs to (1) determine how many subalgorithms it needs to construct and (2) build the body of the main algorithm in parallel. The agent may also need to call subalgorithms from other subalgorithms. In this case, optimizing for length leads to significantly shorter algorithms compared to the human benchmarks as seen in Table 1a. However, owing to the complexities introduced by branching, latency and length are not always correlated; see Supplementary Information for more details. As such, we implemented a procedure that measures the actual latency of the programs by taking the fifth percentile of latency measurements across 100 different machines, with computed confidence intervals[44], and optimize this metric. See Methods for the full benchmarking setup. When optimizing for latency, the agent improves significantly on the human benchmarks in each case as seen in Table 1b.

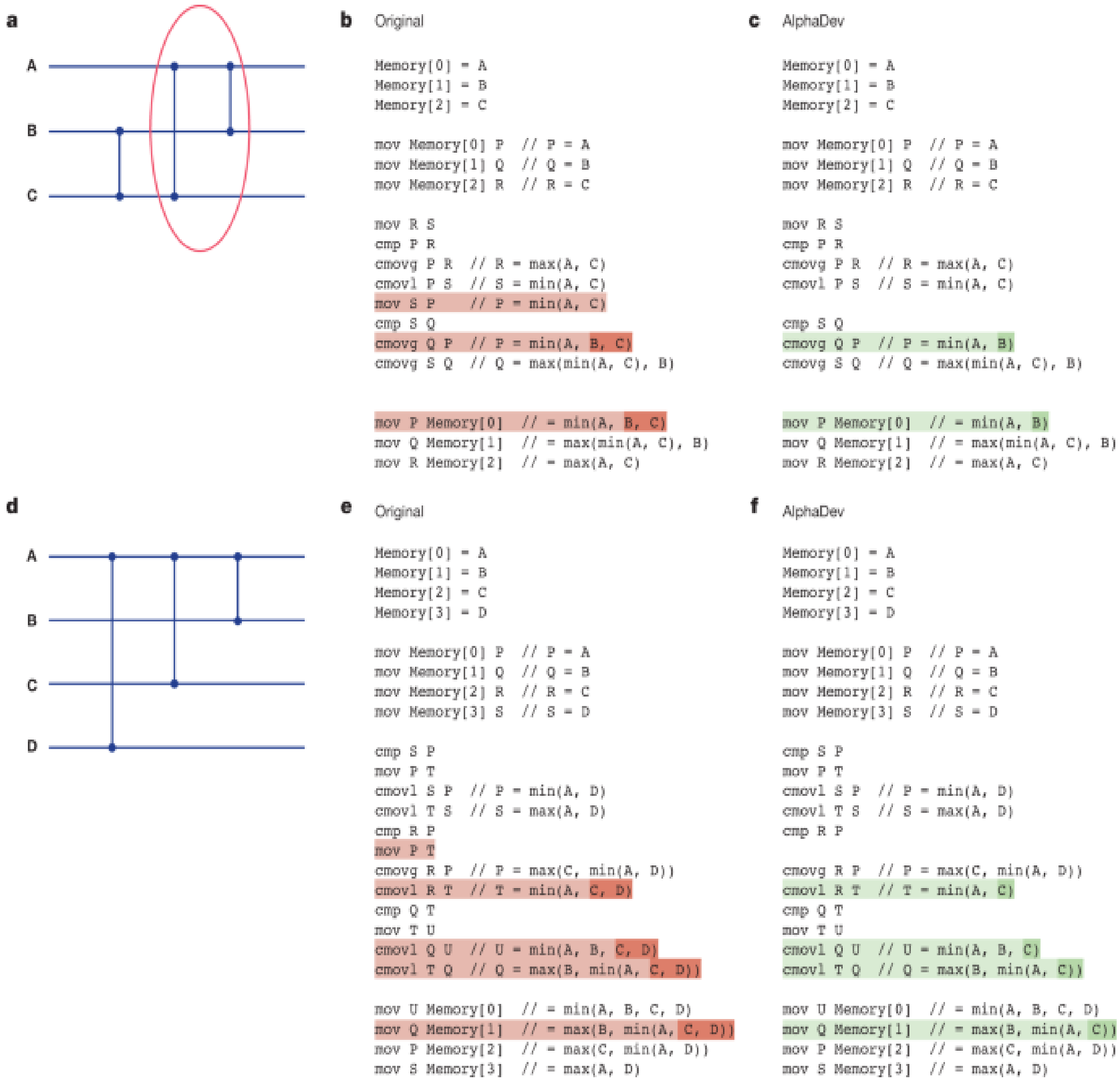
New algorithm discoveries

The solutions discovered by AlphaDev include new and exciting algorithmic discoveries that lead to more efficient performance. In the fixed sort setting, we found that AlphaDev discovered two interesting sequences of instructions that, when applied to a sorting network algorithm, reduce the algorithm by one assembly instruction each time. We refer to each sequence of instructions as (1) the AlphaDev swap move and (2) the AlphaDev copy move respectively.

AlphaDev swap move

Figure 3a presents an optimal sorting network for three elements (see Methods for an overview of sorting networks). We will explain how AlphaDev has improved the circled network segment. There are many variants of this structure that are found in sorting networks of various sizes, and the same argument applies in each case. The circled part of the network (last two comparators) can be seen as a sequence of instructions that takes an input sequence ⟨A, B, C⟩ and transforms each input as shown in Table 2a (left). However, a comparator on wires B and C precedes this operator and therefore input sequences where $B \leq C$ are guaranteed. This means that it is enough to compute min(A, B) as the first output instead of min(A, B, C) as shown in Table 2a (right). The pseudocode difference between Fig. 3b,c demonstrates how the AlphaDev swap move saves one instruction each time it is applied.

Fig. 3: Sorting networks and algorithmic improvements discovered by AlphaDev.



a, An optimal classic sorting network for three inputs. The circled comparators have been improved by AlphaDev. See the AlphaDev swap move for more details. **b**,**c**, The assembly pseudocode before applying the AlphaDev swap move (**b**) and after applying the AlphaDev swap move (**c**), resulting in the removal of a single instruction. **d**, An optimal classic sorting network comparator configuration that has been improved by AlphaDev. See the AlphaDev copy move for more details. **e**,**f**, The assembly pseudocode before applying the AlphaDev copy move (**e**) and after applying the AlphaDev copy move (**f**), resulting in the removal of a single instruction.

Table 2 Analysis of the AlphaDev swap and copy moves

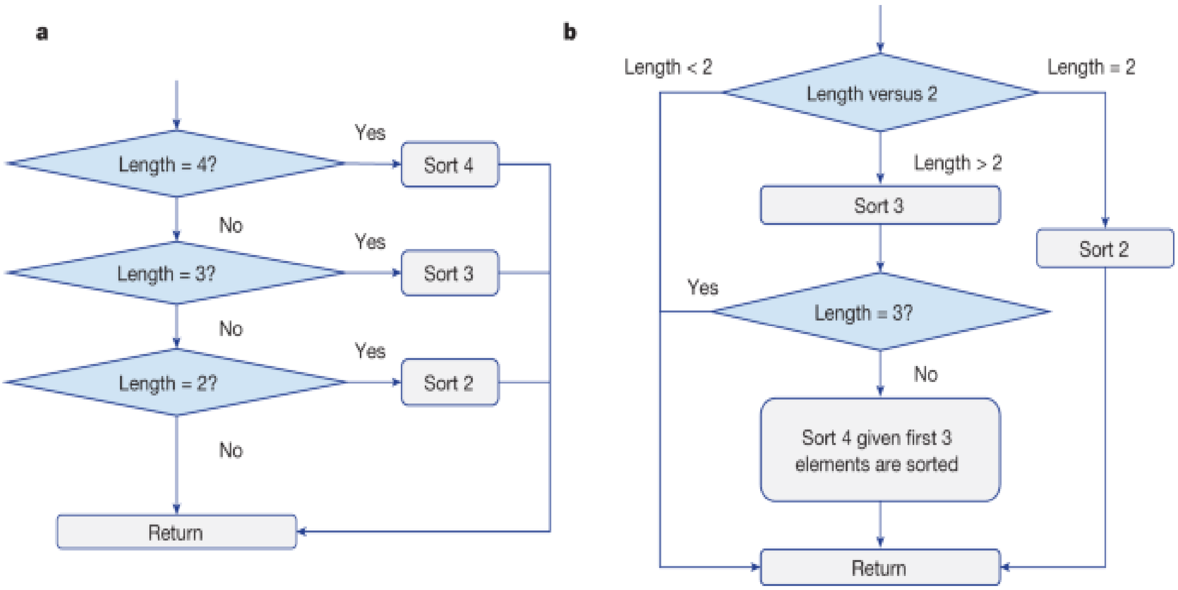
AlphaDev copy move

Figure 3d presents a sorting network configuration, consisting of three comparators, that is applied across four wires. This configuration is found in a sort 8 sorting network and corresponds to an operator taking four inputs $\langle \mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\rangle$ and transforming them into four outputs as seen in Table 2b (on the left). One can show that as part of sort 8, the input that flows into the operator satisfies the following inequality: $\mathrm{D}\ge \min (\mathrm{A}, \mathrm{C})$. This means that the operator can be improved by applying the AlphaDev copy move that is defined in Table 2b (on the right), resulting in one instruction less than the original operator. The code difference between the original operator and the code after applying the AlphaDev copy move is visualized in Fig. 3e,f, respectively.

New variable sort algorithms

The VarSort4 algorithm discovered by AlphaDev is particularly interesting. The flow diagram for the human benchmark algorithm and AlphaDev can be seen in Fig. 4a,b, respectively. The human benchmark algorithm determines the length of the input vector, and then calls the corresponding sorting network to sort the elements. The AlphaDev solution has a completely different approach as seen in Fig. 4b. If the length of the input vector is strictly greater than 2, then sort 3 is immediately called, resulting in the first three elements being sorted. If the vector is greater than three elements, then a simplified sort 4 algorithm is called that sorts the remaining unsorted elements in the input vector. It is this simplified part of the routine that yields significant gains in terms of algorithmic length and latency.

Fig. 4: Fundamentally different algorithms discovered by AlphaDev.



a, A flow diagram of the variable sort 4 (VarSort4) human benchmark algorithm. In this algorithm, a sequence of unsorted numbers are input into the algorithm. If the sequence length is four, three or two numbers, then the corresponding sort 4, sort 3 or sort 2 sorting network is called that sorts the resulting sequence. The result is then returned and output by the function. **b**, The VarSort4 algorithm discovered by AlphaDev. This algorithm also receives sequences of length four, three or two numbers as input. In this case, if the length is two, then it calls the sort 2 sorting network and returns. If the length is three then it calls sort 3 to sort the first three numbers and returns. If, however, the length is greater than three, then it calls sort 3, followed by a simplified sort 4 routine that sorts the remaining unsorted number. It is this part of the routine that results in significant latency savings.

Stochastic search optimization approaches

It is important to understand the advantages and limitations of RL compared to other approaches for program optimization. As such, we implemented a state-of-the-art stochastic superoptimization approach[8], adapted it to the sort setting and used it as the learning algorithm in AlphaDev. We refer to this variant as AlphaDev-S (see Methods for more details). We run this algorithm with at least the same amount of resources and wall-clock time as AlphaDev. AlphaDev-S requires a prohibitive amount of time to optimize directly for latency as latency needs to be computed after every mutation. As

such, AlphaDev-S optimizes for a latency proxy, namely algorithm length and, then, at the end of training, we search through all correct programs generated by AlphaDev-S and benchmark each one to find the lowest latency solution. In general, we find that AlphaDev consistently outperforms AlphaDev-S when learning from scratch without previous knowledge. In addition, as the size of the program increases, AlphaDev explores orders of magnitude fewer programs (12 million programs in the worst case) compared to AlphaDev-S (31 trillion programs in the worst case). This may be because AlphaDev is able to better explore the space of algorithms compared to the breadth-first stochastic search procedure that gets stuck more easily into local optima; see Methods for an overview of this exploration hypothesis. In addition, AlphaDev never evaluates latency during search as it uses the latency value function predictions and, because of this, only needs to compute actual measured latency on less than 0.002% of generated programs. When incorporating previous knowledge into AlphaDev-S, such as warm starting the learning algorithm with a near-optimal solution, AlphaDev-S is more computationally efficient for sort 3, sort 4 and sort 5 (branchless assembly algorithms) and also generates competitive low-latency algorithms to that of AlphaDev in each case. However, for algorithms that require branching (if–else statements), in which algorithm length and latency are not well correlated, AlphaDev discovers lower latency solutions than AlphaDev-S, even when warm starting this algorithm with a near-optimal solution. See Methods for an in-depth analysis of these algorithms.

Generalization to additional domains

To test the generality of AlphaDev, we train the agent on a set of additional domains. These include a protocol buffer deserialization subroutine called VarInt, presented below, and a competitive coding problem (see Appendix D in Supplementary Information for more details). The competitive coding domain latency performance is reported in Table 1b.

Protocol Buffer is Google's open-source data format used to serialize structured data[45]. This format is commonly used in cases in which performance or network load is of primary concern. The VarInt algorithm[46] is a key component in both the serialization and deserialization processes.

We trained the AlphaDev agent as in variable sort to optimize the VarInt deserialization function with respect to correctness and measured latency. For correctness, we reward the agent for correctly deserializing each input. We use a set of 80 inputs and corresponding outputs that cover common protobuf use cases. AlphaDev learns an optimized VarInt deserialization function and manages to significantly outperform the human benchmark for single valued inputs. Our agent discovers a branchless solution that is both shorter (Table 1a) and roughly three times faster than the human benchmark (Table 1b). In doing so, the agent also discovered a new VarInt assignment move in which AlphaDev learns to combine two operations into a single instruction leading to latency savings. See Appendix D.1 in Supplementary Information for a full overview of this move. This is a strong indication that AlphaDev is capable of generalizing to optimize non-trivial, real-world algorithms.

Libc++ sort patch

The sort 3, sort 4 and sort 5 algorithms in the LLVM libc++ standard sorting library are called many times by larger sorting algorithms and are therefore fundamental components of the library. We reverse engineered the low-level assembly sorting algorithms discovered by AlphaDev for sort 3, sort 4 and sort 5 to C++ and discovered that our sort implementations led to improvements of up to 70% for sequences of a length of five and roughly 1.7% for sequences exceeding 250,000 elements. These improvements are for the uint32, uint64 and float data types for ARMv8, Intel Skylake and AMD Zen 2 CPU architectures; see Appendix E in Supplementary Information for the full performance tables. The performance improvements are due to both the branchless conditional assembly generated by AlphaDev as well as the new AlphaDev swap move. For sort 5, we used a 43 length algorithm discovered by AlphaDev, as it led to a more efficient C++ implementation. These algorithms were sent for review and have officially been included in the libc++ standard sorting library[3]. It is the first change to these sub-routines in over a decade. This is also the first time that any component in this sort library has been replaced by an algorithm that has been automatically discovered using reinforcement learning. We estimate that these routines are being called trillions of times every day[1,35,47].

Discussion

AlphaDev discovers new, state-of-the-art sorting algorithms from scratch that have been incorporated into the LLVM C++ library, used by millions of developers and applications around the world[23,24,25]. Both AlphaDev and stochastic search are powerful algorithms. An interesting direction for future research is to investigate combining these algorithms together to realize the complementary advantages of both approaches.

It is important to note that AlphaDev can, in theory, generalize to functions that do not require exhaustive verification of test cases. For example, hashing functions[48] as well as cryptographic hashing functions[49] define function correctness by the number of hashing collisions. Therefore, in this case, AlphaDev can optimize for minimizing collisions as well as latency. AlphaDev can also, in theory, optimize complicated logic components within the body of large, impressive functions. We hope that AlphaDev can provide interesting insights and inspire new approaches in both the artificial intelligence and program synthesis communities.

Methods

Background

AlphaZero

AlphaZero[33] is an RL algorithm that leverages MCTS as a policy improvement operator. It consists of (1) a representation network f^{rep} that outputs a latent representation $\mathbf{h}_t$ of the state $\mathbf{S}_t$; and (2) a prediction network f^{pred} that predicts the expected return (the value) $\hat{v}_{t}$ and a policy (that is, distribution over the action space) $\hat{\pi}_{t}$ from a given latent state. The algorithm uses the true dynamics and reward when planning. MuZero[38] is a model-based variant of AlphaZero that has the same representation and prediction networks, but also learns a model of the dynamics and predicts rewards, which it uses for planning. Specifically, it learns a dynamics network f^{dyn} that predicts the next latent state \

$({\bf{\text{h}}})_{t}^{k+1}$ and reward ${\hat{r}}_{t}^{k+1}$ resulting from a transition. Note that the subscript *t* denotes timesteps in the real environment and the superscript *k* represents timesteps in the model.

$${{\bf{\text{h}}}}_{t}={f}^{rep}({{\bf{\text{S}}}}_{t}) \tag{1}$$

$${{\bf{\text{h}}}}_{t}^{k+1},\,{\hat{r}}_{t}^{k+1}={f}^{dyn}({{\bf{\text{h}}}}_{t}^{k},{{\bf{\text{a}}}}_{t}^{k}) \tag{2}$$

$${\hat{v}}_{t},\,{\hat{\pi }}_{t}={f}^{pred}({{\bf{\text{h}}}}_{t}) \tag{3}$$

On reaching a new state, AlphaZero proceeds by first encoding the state into a latent representation with the representation network. Then, the true dynamics or dynamics network (for MuZero) as well as the prediction network $f^{pred}(\mathbf{h}_t)$ are used to simulate several trajectories that fill out a search tree, by sampling state transitions. At each node, the actions are selected using an optimistic strategy called the predictor upper confidence tree bound[32], meant to balance exploration (trying new actions) and exploitation (progressing further down the subtree of the current estimate of the best action). This strategy starts out by following the predicted policy ${\hat{\pi }}_{t}$ closely, and gradually shifts towards maximizing the predicted value function. Ultimately, an action is recommended by sampling from the root node with probability proportional to its visit count during MCTS. The predicted policy is then trained to match the visit counts of the MCTS policy in an attempt to distil the search procedure into a policy such that subsequent iterations of MCTS will disregard nodes that are not promising.

Sorting networks

Sorting networks are very efficient as their structures can be parallelized on modern CPU architectures. They therefore tend to achieve faster runtime performance, especially on small sorts, compared to popular and efficient base case algorithms such as insertion sort[17,43,50]. A sorting network[43] consists of two types of item called comparators (vertical lines) and wires (horizontal lines) (Extended Data Fig. 2a). Each wire carries a value from left to right. When two wires intersect at a comparator, the values on the two

wires are compared. If the value of the bottom wire is smaller than the value of the top wire, then the values are swapped between wires as seen in Extended Data Fig. 2b. A programmatic implementation of a sorting network consists of executing these swaps on particular pairs of elements from the input sequence in a particular order.

Action pruning rules

We pruned the action space by removing some program invariances (for example, the order of register allocation) and illegal instructions (for example, comparing two memory locations). This helps reducing the size of the action space and increases convergence rate. For our experiments, we used the following rules:

1. (1)

   Memory locations are always read in incremental order.

2. (2)

   Registers are allocated in incremental order.

3. (3)

   We cannot compare or conditionally move to a memory location (illegal).

4. (4)

   We can read and write to each memory location only once.

5. (5)

   We cannot use non-initialized registers (illegal).

6. (6)

   Do not perform consecutive compare instructions.

Training regime

We train AlphaDev on a Tensor Processing Unit (TPU) v.3, with a total batch size of 1,024 per TPU core. We use up to 16 TPU cores and train for 1 million iterations. On the actor side, the games are played on standalone TPU v.4, and we use up to 512 actors. In practice, across all tasks, training takes, in the worst case, 2 days to converge.

AlphaDev-S

It is important to understand the advantages and limitations of RL compared to other possible approaches for program optimization. As such, we implemented a state-of-the-art stochastic superoptimization approach[8] and incorporated it into AlphaDev as the learning algorithm to optimize sorting functions. We refer to this adapted version as AlphaDev-S. Our re-implementation has been specifically optimized for the sorting domain. This includes implementing the algorithm to run with our assembly environment, defining a correctness and performance loss function specific to sorting and running extensive hyperparameter sweeps to identify the best variant. The cost function used for AlphaDev-S is $c = \text{correctness} + \alpha \times \text{performance}$ where correctness corresponds to computing the number of incorrect input sequence elements that are still unsorted, performance corresponds to the algorithm length reward and α is a weight trading off the two cost functions. We are unable to optimize directly for latency as this slows down the learning algorithm considerably making learning infeasible. It should be noted that this function has been adapted to support the same set of assembly instructions used by AlphaDev as well as prune the same set of incorrect or illegal actions. It also uses the same program correctness computation module (Fig. 2b) to compute the correctness term.

AlphaDev-S is then executed by first proposing a transformation to the program stored in the buffer (which may be empty or initialized with an already sorted program). The correctness and performance terms are then computed using the program correctness module and algorithm length, respectively. If the cost is lower than the current best cost, the new program is accepted with high probability, otherwise it is rejected. We will now discuss the correctness cost function and transform weights in more detail.

Correctness cost

For the correctness cost function, we implemented three types of cost function. The first one is defined as the percentage of incorrectly placed items: $\frac{P-P{C}_{t}}{P}$ where P is the total number of items to place and PC_t is number of correctly placed items at timestep t. The second variant is the square root of this equation. The final cost function takes the square root of the difference $\sqrt{-{PC}_{t}}$ and this is what yielded the best performance.

Program transformations

We enabled several program transformations such as adding an instruction to increase the size of the program (Add Transform), swapping two instructions (Swap Transform), randomly changing an Opcode for an instruction (Opcode Transform), randomly sampling an Operand for a chosen instruction (Operand Transform) and randomly sample an Opcode and its corresponding Operands (Instruction Transform). It is possible to influence the sampling of these transforms to encourage some to be sampled more or less frequently. We optimized the weights for sampling transforms by running an extensive hyperparameter sweep.

Investigative studies for AlphaDev variants

We now present a set of investigative studies that help to better understand the advantages and limitations of the DRL and the stochastic search learning algorithms used in AlphaDev. We compare AlphaDev to AlphaDev-S. We implemented two variants of AlphaDev-S: (1) Cold Start (AlphaDev-S-CS) and (2) Warm Start (AlphaDev-S-WS). AlphaDev-S-CS uses no previous information and has to generate a program from an empty program buffer. AlphaDev-S-WS's buffer is warm started with a correct sorting program (for example, optimal sorting network assembly program) and it edits the program to optimize it further. We compared the variants with AlphaDev in both the individual and variable sort algorithm setups.

Because AlphaDev always learns from scratch with no previous knowledge, the direct comparison would be to the cold start stochastic search version:

AlphaDev-S-CS. However, as initial near-optimal programs may sometimes be available, we also compare AlphaDev to the warm start stochastic search version: AlphaDev-S-WS.

It should be noted that the stochastic search variants are unable to optimize directly for latency, as this would make learning infeasible because of computational efficiency. As such, our AlphaDev-S variants optimize for algorithm length. Then, at the end of training, we iterate through the set of generated programs for AlphaDev-S across varying lengths and identify the program with the lowest latency.

In each case, the stochastic search algorithms (AlphaDev-S) are run using at least the same computational resources and wall-clock time to that of AlphaDev.

Fixed sort

We first examine the performance of the various approaches for the fixed sort algorithms. In this case, all algorithmic variants optimize for algorithm length as algorithm length and latency are highly correlated in the conditional branchless setting (see Supplementary Information for more details).

In the cold start setting, AlphaDev-S-CS is unable to find the optimal programs in each case as seen in Extended Data Table 2a. In addition, AlphaDev-S-CS explores orders of magnitude more programs than AlphaDev as shown in Extended Data Table 2b. In the warm start setting, AlphaDev-S is warm started with a near-optimal sorted program, and is able to match the performance of AlphaDev in each case as shown in Extended Data Table 2a. It is more computationally efficient than AlphaDev as shown in Extended Data Table 2c but explores orders of magnitude more programs for sort 3 and sort 5 as shown in Extended Data Table 2b. It can be argued that AlphaDev-S-WS has a substantial advantage in this scenario as it is provided with an initial near-optimal program. We will show in the Variable sort section that when the algorithms become more complicated and branching is introduced, warm starting the learning algorithm with a near-

optimal program is not enough and can cause it to get stuck in suboptimal solutions.

Brute-force approach

We also used a brute-force approach to prove that no program shorter than 17 instructions exists for sort 3. We had to enumerate roughly 10^{32} programs and, even with pruning heuristics, it took more than 3 days to prove this hypothesis. For sort 4 and above this approach is infeasible.

Latency benchmarking suite

The length of a program is only a proxy for the performance of an algorithm. As we introduce branching structures, the length and latency of a program are not well correlated. Therefore, we run the programs on actual machines and measure their latency. Microbenchmarking is very challenging given the numerous noise sources that could affect the measurements. This is especially true when running on shared machines where there could be interference from other processes. Our approach is to have a separate benchmarking service, replicated on separated machines, so that we can quickly perform many measurements in a controlled environment under different conditions. The system works as follows:

1. (1)

   The RL agent processes 1,000 measurements across the machines using the replicated service.

2. (2)

   For each measurement, the service runs the given sorting algorithm over 10,000 random inputs (for example, for sort 3 this would be $3 \times 10{,}000 = 30{,}000$ random integers).

3. (3)

We measure the time taken using a CPU performance counter (CPU_CLK_UNHALTED.CORE).

We then take the fifth percentile as our final measurement, because we assume that most noise sources are one-sided (for example, cache misses, pre-emptions and so on). During training we process the measurements across ten machines for computational efficiency. After training, we benchmark AlphaDev's solution against the baseline solutions, and process the measurements across 100 machines for more accuracy and noise reduction. For each benchmark, we compute confidence intervals using the distribution-free two-sided confidence interval for a quantile tabular method[44].

Variable sort

When optimizing directly for latency, AlphaDev outperforms AlphaDev-S-WS on VarSort3, VarSort4 and VarSort5 as seen in Extended Data Table 3a. AlphaDev-S-CS fails to find a solution in each case. In the cases of VarSort4 and VarSort5, program length and latency are not correlated (see Supplementary Information for more details). This indicates that when program length cannot be used as a proxy for performance, AlphaDev is able to find lower latency solutions compared to AlphaDev-S. This is even in the case where the stochastic search is warm started with a near-optimal program. In addition, AlphaDev converges to the optimal solution after exploring a maximum of 12M programs as seen in Extended Data Table 3b. This is orders of magnitude lower than that of AlphaDev-S-CS and AlphaDev-S-WS, respectively (31 trillion programs in the worst case).

Exploration hypothesis

We proposed that AlphaDev-S struggles to discover programs when learning from scratch and gets stuck in local optima when warm started because of its limited exploration capabilities as a result of the stochastic search procedure. Extended Data Fig. 3 shows two-dimensional *t*-stochastic neighbour embedding (*t*-SNE) projections[51] of AlphaDev and AlphaDev-S's assembly algorithms discovered during their respective training procedures for VarSort5. The features used in the projection include correctness, latency,

algorithm length and a histogram count of the instructions used per algorithm. Extended Data Fig. 3a indicates the regions in algorithm space explored by AlphaDev, AlphaDev-S-CS and AlphaDev-S-WS, respectively, whereas Extended Data Fig. 3b superimposes algorithm correctness onto each point in the *t*-SNE projection in which the colour indicates the correctness of each discovered algorithm, ranging from incorrect algorithms (purple) to correct algorithms (yellow). The AlphaDev-S variants both cover a densely packed circular region around their initial seed, which highlights the breadth-first nature of their stochastic search procedure. This illustrates that AlphaDev-S-CS fails to navigate through the space of incorrect algorithms in a reasonable amount of time and discover correct algorithms when learning from scratch. A similar argument applies to AlphaDev-S-WS whereby, when optimizing from an already correct but suboptimal expert demonstration, the algorithm is biased towards exploring its vicinity and struggles to escape this local maxima. By contrast, AlphaDev has more diverse algorithm space coverage, as the long-term value function is a guiding signal for discovering new and interesting parts of algorithm space. As seen in Extended Data Fig. 3b, it is capable of escaping the space of incorrect algorithms to discover a new space of correct algorithms, highlighting the exploration advantages afforded by AlphaDev.

Related work

Assembly optimization

There are numerous approaches to optimizing assembly programs, which we have classified into three groups: enumerative search, stochastic search and symbolic search[5].

First, enumerative search techniques include brute-force program enumeration[4,5,6] as well as implicit enumeration using symbolic theorem proving[52,53]. These approaches search through the space of programs to find a solution based on a predefined set of programs, heuristic and/or cost function. These approaches struggle to span large regions of program space, especially as the size and complexity of the program increases.

Second, stochastic search techniques circumvent comprehensive enumeration by relying on sampling mechanisms such as Markov chain Monte Carlo sampling[5,6,8,9]. Rajeev Alur et al.[5] define a correctness specification, provided by a logical formula that uses symbols from a background theory. The goal is to then find an implementation expression such that logical formula defining the specification is valid. The idea is to iteratively add test cases and then search and expand the program to solve the given test cases. They optimize for correctness on problems from the book Hacker's delight[54]. Phitchaya Mangpo Phothilimthana et al.[6] introduce the LENS algorithm that is based on running enumerative, stochastic and symbolic search in parallel, while relying on handcrafted pruning rules. This setup is capable of optimizing up to 21 instructions, and cannot optimize for latency nor support branching. Another algorithm[8] is based on Markov chain Monte Carlo rejection sampling and applies transformations to programs in assembly using a loss function that is a function of correctness and performance. Many of these approaches are prone to getting stuck in local minima and may also struggle as the size and/or complexity of the program increases. In addition, incorporating actual, measured latency into these approaches are either infeasible or prohibitively expensive.

Third, symbolic search approaches can also be implemented to optimize assembly programs. These include SAT solvers[55], SMT solvers[5,6] and Mixed Integer Programs (MIPs)[56,57]. However, these approaches suffer from scaling issues. For example, classical solvers require a problem to be translated into a certain canonical form. It usually requires an expert in the said solvers and a substantial amount of time to find an efficient formulation. In addition, for any new modification of the problem, this has to be repeated. Classical solvers are also hard to parallelize and thus, it is challenging to leverage more hardware to speed up the solving process. Another symbolic search algorithm is Cholorphyll[10] that implements a multi-phase approach. It first requires as input a source program with partition annotations that specify where code and data reside. Then, a layout synthesizer maps program fragments onto physical cores to minimize computational costs. The code is then separated into per-core program fragments and the program fragments are compiled into machine code. At this point, a superoptimizer optimizes each of these fragments.

SIMD optimization

Various approaches[58,59,60] have also been applied to sorting functions that run in the single instruction, multiple data (SIMD)[61] setup. This setup is capable of parallelizing instruction execution, but is not supported at present in popular libraries such as LLVM's libc++ std::sort library. One example is that from Gilles Barthe et al.[7] that proposes a methodology for optimizing programs by automatically vectorizing loops with SIMD instructions. They do this by introducing a framework for verifying the correctness of transformations to a program and performing a search-based procedure using the said transformation. Their framework can discover SIMD looping structures of up to nine instructions in 0.12 s, which corresponds to a minimum 2× speed-up.

RL approaches for program synthesis

There are also several studies using RL for program optimization. Kevin Ellis et al.[62] learn a policy and value function to write and evaluate code, as well as performing a Monte Carlo-style search strategy during inference. This work requires a pretraining step and aims to generate correct programs that satisfy a predefined specification. The approach is successfully applied to computer-aided design and string editing programs. SuperSonic[63] uses an RL meta-optimizer to select between different RL architectures, using a Multi-Armed Bandit policy search to find a state representation, reward function and RL algorithm that is optimal for the current task. This requires keeping track of many RL algorithms and architectures, which are used as part of the state space. By contrast, our approach only focuses on training a single RL architecture, taking advantage of MCTS search and powerful state representations. Shypula et al.[64] create a supervised assembly dataset and use it to train a Transformer model for mapping unoptimized to optimized code, followed by an RL stage for improving the solution quality. Our method does not require a supervised dataset or two separate training and finetuning stages, and optimizes everything end-to-end using RL and search instead. Chen et al.[65] define their own domain specific language and perform input–output program synthesis that better uses the intermediate program representation to guide the synthesis routine. They show that this can be

incorporated with RL, using the setup of Rudy Bunel et al.[66] and improve the correctness of generated functions. They do not, however, optimize for program length or latency.

Input–output examples for program synthesis

A large body of work addresses the problem of learning programs from input–output pairs. One type of approach learns a neural network for matching inputs to outputs directly[11,13,67,68]. This approach is difficult to integrate into existing libraries and can struggle to generalize to previously unseen inputs, although there has been some encouraging recent progress using graph representations[69]. Another type of approach is to perform a search in program space, guided by a learned model[12,70,71,72]. For instance, Chen et al.[70] use a model that predicts the next program token on the basis of a partial program and the input–output pairs. This bears some similarities to how search is guided in our approach: the learned policy prior in AlphaZero is a model for predicting the next token, learned on the basis of a combination of a partial program and that program's effects on the inputs. However, we are interested in finding correct and efficient programs, which we achieve by further learning a value function for approximating the expected latency of partial programs, and using AlphaZero to incorporate this value function into the search process.

Deep learning for code generation

There are also several deep learning approaches that use large languages models to generate code. These approaches vary in their uses from transpilation, code refactoring and explaining code[15] to generating human-level competitive code using a natural language description[14]. That particular work aims to generate correct code, but does not focus on generating low-latency solutions.

Sort-based program optimization

There are several program synthesis studies that have tackled sorting algorithms. For example, White et al.[26] use RL for learning sorting

functions. Their work uses several heuristics and a domain specific language to yield a sorting algorithm called reinforcement programming sort. Srivastava et al.[27] encodes the program synthesis as a verification problem. Specifically, they represent a synthesis task as a tuple consisting of the functional expression, the domains and guards appearing in the synthesized program and the resource constraints. The idea is that, given a prespecified resource constraint, their synthesizer produces a program that meets the predefined specification to ensure correctness. They apply this to discover merge sort and quick sort. Jason Ansel et al.[28] takes as input predefined algorithms (for example, insertion sort, merge sort and quick sort) and then determines when to select these algorithms for execution using its autotuner function. It does so by defining a language that contains rules and transforms that dictate how the algorithms are selected and where they are executed.

Data availability

The data used to train the system were generated synthetically according to the procedures explained in the paper. The algorithms discovered by AlphaDev for the copy and swap operators are presented in the main paper. We have also released the discovered AlphaDev assembly implementations for sort 3–8 as well as VarSort3, 4 and 5 on Github at https://github.com/deepmind/alphadev. We have included exhaustive tests to ensure that each implementation is correct. In addition, Appendix G in Supplementary Information contains a list of additional, correct sorting algorithms discovered by AlphaDev for sort 3, sort 4 and sort 5. The performance of the sort 3, sort 4 and sort 5 algorithms on the official LLVM benchmarking suite for three different CPU architectures as well as floats, int32 and int64 data types is detailed in Appendix E in the Supplementary Information. In addition, the AlphaDev sort 3, sort 4 and sort 5 implementations can be found in the LLVM libc++ standard sorting library[3].

Code availability

We have also released pseudocode at https://github.com/deepmind/alphadev that includes the environment, the full actor and training loops as well as the core MCTS algorithm. In addition, we include our actual JAX

implementation of our policy, value and representation networks that enable the architectures to be reproduced. Finally, we have a config file containing the hyperparameter definitions to be used with the agent.

References

1. Amazon. Amazon S3—two trillion objects, 1.1 million requests/second. *AWS* https://aws.amazon.com/blogs/aws/amazon-s3-two-trillion-objects-11-million-requests-second/ (2013).

2. Cormen, T. H. et al. *Introduction to Algorithms* (MIT Press, 2022).

3. Gelmi, M. Introduce branchless sorting functions for sort3, sort4 and sort5. *LLVM.org* https://reviews.llvm.org/D118029 (2022).

4. Bansal, S. & Aiken, A. Automatic generation of peephole superoptimizers. *ACM SIGARCH Comput. Arch. News* **34**, 394–403 (2006).

5. Alur, R. et al. *Syntax-Guided Synthesis* (IEEE, 2013).

6. Phothilimthana, P. M. et al. Scaling up superoptimization. In *Proc. Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems* 297–310 (ACM, 2016).

7. Barthe, G. et al. From relational verification to SIMD loop synthesis. In *Proc. of the 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* 123–134 (ACM, 2013).

8. Schkufza, E., Sharma, R. & Aiken, A. Stochastic superoptimization. *ACM SIGPLAN Notices* **48**, 305–315 (2013).

9. Bunel, R. et al. Learning to superoptimize programs. In *Proc. International Conference on Learning Representations* (ICLR, 2016).

10. Phothilimthana, P. M. et al. Chlorophyll: synthesis-aided compiler for low-power spatial architectures. *ACM SIGPLAN Notices* **49**, 396–407

(2014).

11. Vinyals, O. et al. Grammar as a foreign language. *Adv. Neural Inform. Proc. Syst.* **28**, 2773–2781 (2015).

12. Chen, X., Liu, C. & Song, D. Towards synthesizing complex programs from input-output examples. In *Proc. International Conference on Learning Representations* (ICLR, 2018).

13. Devlin, J. et al. Robustfill: neural program learning under noisy i/o. In *Proc. International Conference on Machine Learning* 990–998 (PMLR, 2017).

14. Li, Y. et al. Competition-level code generation with AlphaCode. *Science* **378**, 1092–1097 (2022).

15. Pearce, H. et al. Can codex and other large language models help us fix security bugs? Preprint at https://arxiv.org/abs/2112.02125 (2021).

16. Chen, M. et al. Evaluating large language models trained on code. Preprint at https://arxiv.org/abs/2107.03374 (2021).

17. Bingmann, T., Marianczuk, J. & Sanders, P. Engineering faster sorters for small sets of items. *Software: Pract. Exper.* **51**, 965–1004 (2021).

18. Levcopoulos, C. & Petersson, O. Splitsort: an adaptive sorting algorithm. *Inform. Proc. Lett.* **39**, 205–211 (1991).

19. Helman, D. R., Bader, D. A. & JáJá, J. A randomized parallel sorting algorithm with an experimental study. *J. Parallel Distrib. Comput.* **52**, 1–23 (1998).

20. Goodrich, M. T. Randomized shellsort: a simple oblivious sorting algorithm. In *Proc. of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* 1262–1277 (ACM, 2010).

21. Mehlhorn, K., Sanders, P. & Sanders, P. *Algorithms and Data Structures: The Basic Toolbox* Vol. 55. (Springer, 2008).

22. Knebl, H. *Algorithms and Data Structures* (Springer, 2020).

23. Karatzoglou, A., Baltrunas, L. & Shi, Y. Learning to rank for recommender systems. In *Proc. of the 7th ACM Conference on Recommender Systems* 493–494 (ACM, 2013).

24. Yang, J. Y., Zhang, B. & Mao, Y. Study on Information Retrieval Sorting Algorithm in Network-BasedManufacturing Environment. In *Applied Mechanics and Materials* Vol. 484, 183–186 (Trans Tech Publishing, 2014).

25. Krallmann, J., Schwiegelshohn, U. & Yahyapour, R. On the design and evaluation of job schedulingalgorithms. In *Workshop on Job Scheduling Strategies for Parallel Processing* 17–42 (Springer, 1999).

26. White, S. K., Martinez, T. & Rudolph, G. Generating a novel sort algorithm using Reinforcement Programming. In *Proc. IEEE Congress on Evolutionary Computation* 1–8 (IEEE, 2010).

27. Srivastava, S., Gulwani, S. & Foster, J. S. From program verification to program synthesis. In *Proc. of the 37th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* 313–326 (ACM, 2010).

28. Ansel, J. et al. Petabricks: a language and compiler for algorithmic choice. *ACM Sigplan Notices* **44**, 38–49 (2009).

29. Smith, D. R. The design of divide and conquer algorithms. *Sci. Comput. Program.* **5**, 37–58 (1985).

30. Irvine, K. R. et al. *Assembly Language for Intel-Based Computers* (Prentice Hall, 2003).

31. Shannon, C. E. XXII. Programming a computer for playing chess. *London, Edinb. Dublin Philos. Mag. J. Sci.* **41.314**, 256–275 (1950).

32. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

33. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).

34. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inform. Proc. Syst.* **30**, 5999–6009 (2017).

35. LLVM. LLVM users https://llvm.org/Users.html (LLVM, 2022).

36. Bartlett, J. *Learn to Program with Assembly* 271–273 (Apress, 2021).

37. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (MIT Press, 2018).

38. Schrittwieser, J. et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature* **588**, 604–609 (2020).

39. Maillard, O.-A., Ryabko, D. & Munos, R. Selecting the state-representation in reinforcement learning. *Adv. Neural Inform. Proc. Syst.* **24**, 2627–2635 (2011).

40. Qian, R. et al. Spatiotemporal contrastive video representation learning. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 6964–6974 (IEEE, 2021).

41. Brown, T. et al. Language models are few-shot learners. *Adv. Neural Inform. Proc. Syst.* **33**, 1877–1901 (2020).

42. Shazeer, N. Fast transformer decoding: one write-head is all you need. Preprint at https://arxiv.org/abs/1911.02150 (2019).

43. Bundala, D. & Závodny, J. Optimal sorting networks. In *Proc. International Conference on Language and Automata Theory and Applications* 236–247 (Springer, 2014).

44. Hahn, G. J. & Meeker, W. Q. *Statistical Intervals: A Guide for Practitioners* Vol. 92 (John Wiley & Sons, 2011).

45. Google. Protocol buffers, version 0.2.5; https://developers.google.com/protocol-buffers (2022).

46. Google. VarInt protocol buffer serialization and deserialization, version 0.2.5; https://developers.google.com/protocol-buffers/docs/encoding (2022).

47. Protvin, R. & Levenberg, J. Why Google stores billions of lines of code in a single repository. *Commun. ACM* **59**, 78–87 (2016).

48. Berman, I. et al. Multi-collision resistant hash functions and their applications. In *Proc. Annual International Conference on the Theory and Applications of Cryptographic Techniques* 133–161 (Springer, 2018).

49. Damgård, I. B. Collision free hash functions and public key signature schemes. In *Workshop on the Theory and Application of of Cryptographic Techniques* 203–216 (Springer, 1987).

50. Hwang, M. Sort, Bitset (GitHub, 2021).

51. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9.11**, 2579–2605 (2008).

52. Gulwani, S. et al. Synthesis of loop-free programs. *ACM SIGPLAN Notices* **46.6**, 62–73 (2011).

53. Sasnauskas, R. et al. Souper: a synthesizing superoptimizer. Preprint at https://arxiv.org/abs/1711.04422 (2017).

54. Warren, H. S. *Hacker's Delight* (Pearson Education, 2013).

55. Hamadi, Y., Jabbour, S. & Sais, L. ManySAT: a parallel SAT solver. *J. Satisfiability, Boolean Model. Comput.* **6**, 245–262 (2010).

56. Wolsey, L. A. Mixed integer programming. In *Wiley Encyclopedia of Computer Science and Engineering* 1–10 (Wiley, 2007).

57. Nair, V. et al. Solving mixed integer programs using neural networks. Preprint at https://arxiv.org/abs/2012.13349 (2020).

58. Inoue, H. et al. AA-sort: a new parallel sorting algorithm for multi-core SIMD processors. In *Proc. International Conference on Parallel Architecture and Compilation Techniques (PACT 2007)* 189–198 (IEEE, 2007).

59. Yin, Z. et al. Efficient parallel sort on avx-512-based multi-core and many-core architectures. In *Proc. IEEE 21st International Conference on High Performance Computing and Communications* 168–176 (IEEE, 2019).

60. Blacher, M. et al. Vectorized and performance-portable Quicksort. Preprint at https://arxiv.org/abs/2205.05982 (2022).

61. Wikipedia. Single instruction, multiple data https://en.m.wikipedia.org/wiki/SIMD (2022).

62. Ellis, K. et al. Write, execute, assess: program synthesis with a REPL. *Adv. Neural Inform. Proc. Syst.***32**, 9137–9146 (2019).

63. Wang, H. et al. Automating reinforcement learning architecture design for code optimization. In *Proc. 31st ACM SIGPLAN International Conference on Compiler Construction* 129–143 (ACM, 2022).

64. Shypula, A. G. et al. Learning to superoptimize real-world programs. Preprint at https://arxiv.org/abs/2109.13498 (2022).

65. Chen, X., Liu, C. & Song, D. Execution-guided neural program synthesis. In *Proc. International Conference on Learning Representations* (ICLR, 2018).

66. Bunel, R. et al. Leveraging grammar and reinforcement learning for neural program synthesis. In *Proc. International Conference on Learning Representations* (ICLR, 2018).

67. Aharoni, R. & Goldberg, Y. Towards string-to-tree neural machine translation. In *Proc. 55th Annual Meeting of the Association for Computational Linguistics*132–140 (ACL, 2017).

68. Dong, L. & Lapata, M. Language to logical form with neural attention. In *Proc. 54th Annual Meeting of the Association for Computational Linguistics* 33–43 (ACL, 2016).

69. Ibarz, B. et al. A generalist neural algorithmic learner. In *Proc. Learning on Graphs Conference* Vol. 198, 2:1–2:23 (PMLR, 2022).

70. Chen, X., Song, D. & Tian, Y. Latent execution for neural program synthesis beyond domain-specific languages. *Adv. Neural Inform. Proc. Syst.* **34**, 22196–22208 (2021).

71. Parisotto, E. et al. Neuro-symbolic program synthesis. Preprint at https://arxiv.org/abs/1611.01855 (2016).

72. Ellis, K., Solar-Lezama, A. & Tenenbaum, J. Sampling for Bayesian program learning. *Adv. Neural Inform. Proc. Syst.* **29**, 1297–1305 (2016).

Acknowledgements

We thank P. Kurylowicz, N. Anderson and Z. Ahmed for assistance coordinating the research; L. Dionne and N. Klauser for patiently reviewing our LLVM code; and N. Vaish, D. Gove, D. Kutenin and A. Fawzi for their helpful advice during the course of the project. We also thank our colleagues at DeepMind for their encouragement and support.

Author information

Author notes

1. These authors contributed equally: Daniel J. Mankowitz, Andrea Michi, Anton Zhernov, Marco Gelmi, Marco Selvi, Cosmin Paduraru, Edouard Leurent

Authors and Affiliations

1. Deepmind, London, UK

Daniel J. Mankowitz, Andrea Michi, Anton Zhernov, Marco Gelmi, Marco Selvi, Cosmin Paduraru, Edouard Leurent, Shariq Iqbal, Jean-Baptiste Lespiau, Alex Ahern, Thomas Köppe, Kevin Millikin, Stephen Gaffney, Sophie Elster, Jackson Broshear, Chris Gamble, Kieran Milan, Robert Tung, Taylan Cemgil, Mohammadamin Barekatain, Yujia Li, Amol Mandhane, Thomas Hubert, Julian Schrittwieser, Demis Hassabis, Pushmeet Kohli, Martin Riedmiller, Oriol Vinyals & David Silver

2. Google, Mountain View, CA, USA

Minjae Hwang

Contributions

D.J.M., A.Michi and A.Z. conceived the idea and lead the research. A.Michi, D.J.M., A.Z., M.G., M.S., C.P., E.L., S.I. and A.Mandhane developed the neural network architecture and training. J.-B.L., C.P., M.G., D.J.M. and E.L. developed the baseline. M.G., A.Z., D.J.M., M.H., A.A., T.K. and K.Millikin analysed the generated algorithms and helped with the sort patch. D.J.M., A.Michi, A.Z., S.G., S.E., J.B., R.T., C.G. and K.Milan, managed the research. A.Michi, M.G and M.S. led the technical platform. A.Mandhane, T.H., Y.L., J.S., T.C., M.B., P.K., M.R., D.S., O.V. and D.H. contributed technical advice and ideas. D.J.M. and A.Z. conceived the project. D.J.M., C.P., E.L., A.Michi, M.G., A.Z., P.K. and M.S. wrote the paper.

Corresponding author

Correspondence to Daniel J. Mankowitz.

Ethics declarations

Competing interests

D.J.M., A.Michi, A.Z., M.G., M.S., C.P., E.L., S.I., A.Mandhane, P.K., M.R., D.S. and O.V. are planning to file a patent application relating to

subject matter contained in this paper in the name of DeepMind Technologies Limited. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks Zheng Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 The AlphaDev representation network architecture.

(a) The AlphaDev representation network comprises a Transformer Encoder network that receives as input the assembly algorithm generated thus far. It also contains a CPU State Encoder network that receives as input the current state of memory and registers. The exact architecture and hyperparameters can be found in the Supplementary Information, Appendix A. **(b)** Before inputting instructions into the Transformer Encoder network, each program instruction's opcode and operands are converted to one-hot encodings and concatenated. The resulting encoding is then fed into the Transformer Encoder network.

Extended Data Fig. 2 An example sorting network43.

(a) The horizontal lines are called wires and the vertical lines are called comparators. **(b)** An initially unsorted sequence of values are input into the

sorting network on the left hand side. At various stages two wires encounter a comparator. If the value at the top of the comparator is smaller than the value at the bottom of the comparator, the numbers switch wires. An optimal sorting network places comparators in specific positions so as to sort any sequence of unsorted values using the minimum number of comparators.

Extended Data Fig. 3 Hypothesis for improved exploration using AlphaDev.

(a) A 2D t-SNE[51] projection indicating the regions explored by AlphaDev (blue) compared to AlphaDev-S. **(b)** The same 2D t-SNE projection as in (a) with algorithm correctness superimposed onto each point from incorrect programs (purple) to correct programs (yellow). As seen in the figure, AlphaDev-S struggles to move out of local optima whereas AlphaDev is able to explore from the space of incorrect programs to the space of correct programs.

Extended Data Table 1 Additional Assembly instructions
Extended Data Table 2 Comparison of AlphaDev and AlphaDev-S for fixed sort
Extended Data Table 3 Comparison of AlphaDev and AlphaDev-S on variable sort

Supplementary information

Supplementary Information

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Mankowitz, D.J., Michi, A., Zhernov, A. *et al.* Faster sorting algorithms discovered using deep reinforcement learning. *Nature* **618**, 257–263 (2023). https://doi.org/10.1038/s41586-023-06004-9

- Received: 25 July 2022
- Accepted: 23 March 2023
- Published: 07 June 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06004-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

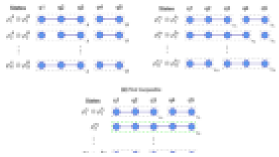
Provided by the Springer Nature SharedIt content-sharing initiative

AI learns to write sorting software on its own

- Armando Solar-Lezama

Nature News & Views 07 Jun 2023

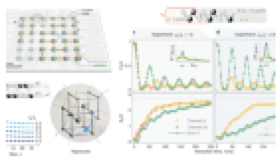
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06004-9



Entanglement in a 20-Qubit Superconducting Quantum Computer

17 September 2019

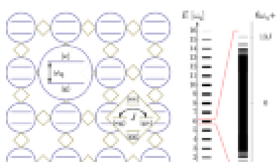
Gary J. Mooney, Charles D. Hill & Lloyd C. L. Hollenberg



Many-body Hilbert space scarring on a superconducting processor

13 October 2022

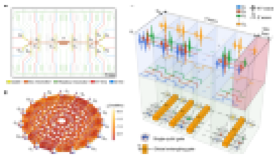
Pengfei Zhang, Hang Dong, … Ying-Cheng Lai



Two-dimensional hard-core Bose–Hubbard model with superconducting qubits

26 June 2020

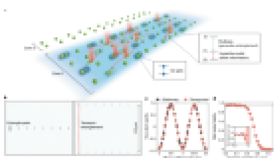
Yariv Yanay, Jochen Braumüller, … Charles Tahan



Observation of entanglement transition of pseudo-random mixed states

08 April 2023

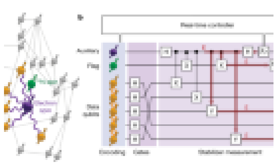
Tong Liu, Shang Liu, … Heng Fan



A quantum processor based on coherent transport of entangled atom arrays

20 April 2022

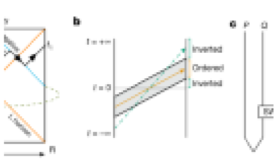
Dolev Bluvstein, Harry Levine, … Mikhail D. Lukin



Fault-tolerant operation of a logical qubit in a diamond quantum processor

05 May 2022

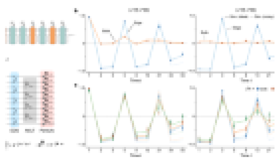
M. H. Abobeih, Y. Wang, … T. H. Taminiau



[Traversable wormhole dynamics on a quantum processor]()

30 November 2022

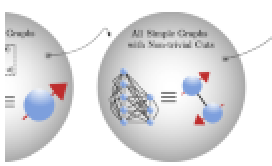
Daniel Jafferis, Alexander Zlokapa, … Maria Spiropulu



[Dynamical topological phase realized in a trapped-ion quantum simulator]()

20 July 2022

Philipp T. Dumitrescu, Justin G. Bohnet, … Andrew C. Potter



[Quantum physics in connected worlds]()

02 December 2022

Joseph Tindall, Amy Searle, … Dieter Jaksch

- Article
- [Open Access]()
- [Published: 11 May 2023]()

Non-Abelian braiding of graph vertices in a superconducting processor

- Google Quantum AI and Collaborators

Nature **volume 618**, pages 264–269 (2023)

- 19k Accesses
- 466 Altmetric
- Metrics details

Abstract

Indistinguishability of particles is a fundamental principle of quantum mechanics[1]. For all elementary and quasiparticles observed to date—including fermions, bosons and Abelian anyons—this principle guarantees that the braiding of identical particles leaves the system unchanged[2,3]. However, in two spatial dimensions, an intriguing possibility exists: braiding of non-Abelian anyons causes rotations in a space of topologically degenerate wavefunctions[4,5,6,7,8]. Hence, it can change the observables of the system without violating the principle of indistinguishability. Despite the well-developed mathematical description of non-Abelian anyons and numerous theoretical proposals[9,10,11,12,13,14,15,16,17,18,19,20,21,22], the experimental observation of their exchange statistics has remained elusive for decades. Controllable many-body quantum states generated on quantum processors offer another path for exploring these fundamental phenomena. Whereas efforts on conventional solid-state platforms typically involve Hamiltonian dynamics of quasiparticles, superconducting quantum processors allow for directly manipulating the many-body wavefunction by means of unitary gates. Building on predictions that stabilizer codes can host projective non-Abelian Ising anyons[9,10], we implement a generalized stabilizer code and unitary protocol[23] to create and braid them. This allows us to experimentally verify the fusion rules of the anyons and braid them to realize their statistics. We then study the prospect of using the anyons for quantum computation and use braiding to create an entangled state of anyons encoding three logical qubits. Our work provides new insights about non-Abelian braiding and, through the future inclusion of error correction to

achieve topological protection, could open a path towards fault-tolerant quantum computing.

Main

Elementary particles in three dimensions are either bosons or fermions. The existence of only two types is rooted in the fact that the worldlines of two particles in three plus one dimensions can always be untied in a trivial manner. Hence, exchanging a pair of indistinguishable particles twice is topologically equivalent to not exchanging them at all, and the wavefunction must remain the same. Representing the exchange as a matrix R acting on the space of wavefunctions with a constant number of particles, it is thus required that $R^2 = 1$ (a scalar), leaving two possibilities: $R = 1$ (bosons) and $R = -1$ (fermions). Such continuous deformation is not possible in two dimensions, thus allowing collective excitations (quasiparticles) to show richer braiding behaviour. In particular, this permits the existence of Abelian anyons[2,3,6,7,8,24,25], in which the global phase change due to braiding can take any value. It has been proposed that there exists another class of quasiparticles known as non-Abelian anyons, in which braiding instead results in a change of the observables of the wavefunction[4,5,24]. In other words, R^2 does not simplify to a scalar, but remains a unitary matrix. Therefore, R^2 is a fundamental characteristic of anyon braiding. The topological approach to quantum computation[26] aims to leverage these non-Abelian anyons and their topological nature to enable gate operations that are protected against local perturbations and decoherence errors[5,27,28,29,30]. In solid-state systems, primary candidates of non-Abelian quasiparticles are low-energy excitations in Hamiltonian systems, including the 5/2 fractional quantum Hall states[31,32], vortices in topological superconductors[33,34] and Majorana zero modes in semiconductors proximitized by superconductors[35,36,37,38]. However, direct verification of non-Abelian exchange statistics has remained elusive[39,40,41].

We formulate the necessary requirements for experimentally certifying a physical system as a platform for topological quantum computation[5,26]: (1) create an anyon pair; (2) verify the rules that govern the 'collision' of two anyons, known as the fusion rules; (3) verify the non-Abelian braiding

statistics reflected in the matrix structure R^2 and (4) realize controlled entanglement of anyonic degrees of freedom. Notably, the observation of steps (2)–(4) requires measurements of multi-anyon states, by means of fusion or non-local measurements.

The advent of quantum processors allows for controlled unitary evolution and direct access to the wavefunction rather than the parameters of the Hamiltonian. These features enable the use of local operations for efficient preparation of topological states that can host non-Abelian anyons, and—as we will demonstrate—their subsequent braiding and fusion. Moreover, these platforms allow for probing arbitrary Pauli strings through destructive multiqubit (that is, non-local) measurements. As the braiding of non-Abelian anyons in this platform is achieved through unitary gate control rather than adiabatic evolution of a Hamiltonian system, we note that the anyons are not quasiparticles in the sense of eigenstates that persist throughout a Hamiltonian evolution. Their movement is achieved through local operations along their paths, and they are kept spatially separated throughout the braiding. We therefore emphasize that the two-dimensional braiding processes are physically taking place on the device, leading to actual non-Abelian exchange effects of local anyons in the many-body wavefunction, rather than matrix operations that simply follow the same algebra.

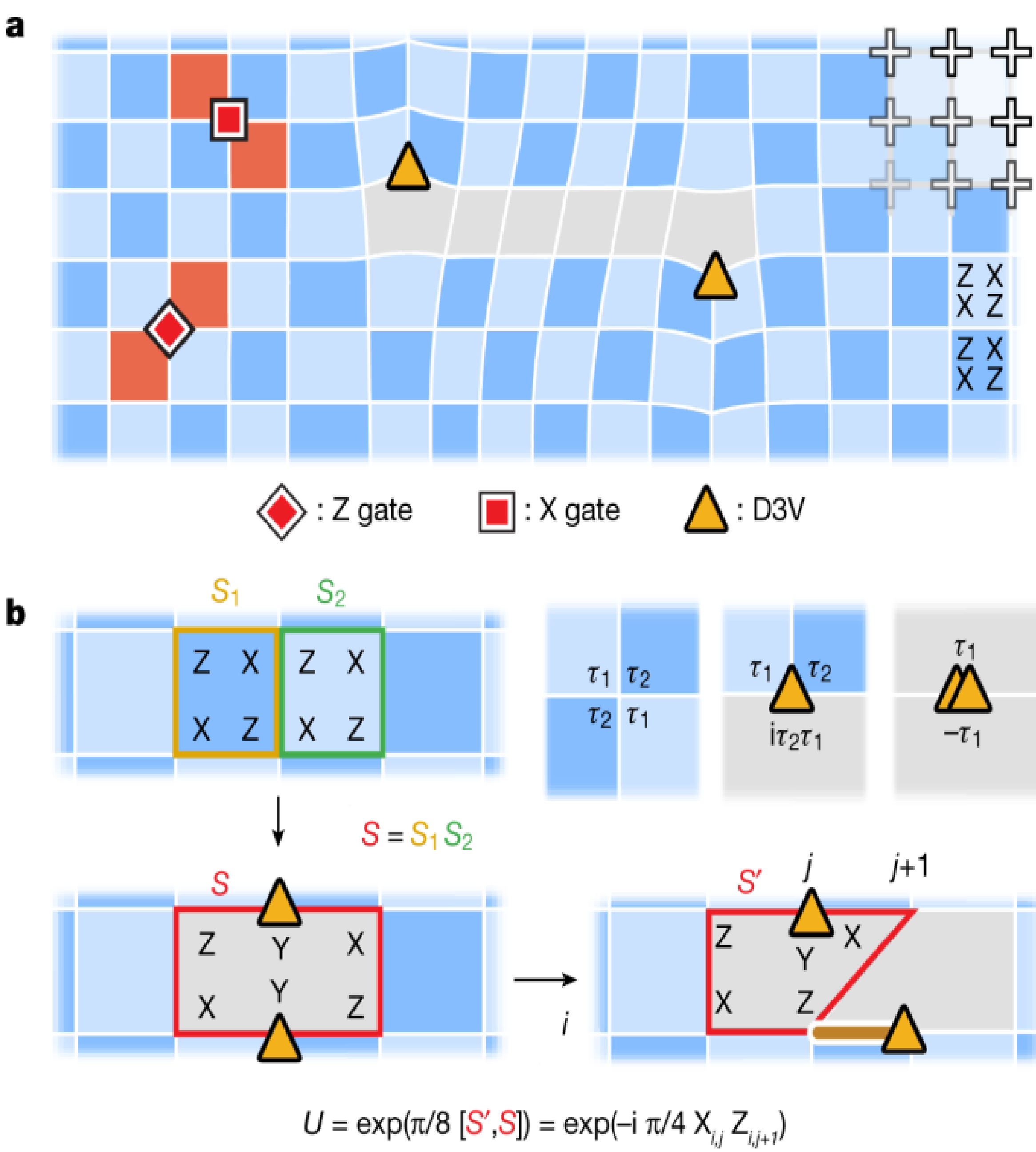
To realize a many-body quantum state that can host anyons, it is essential to control the topological degeneracy. A suitable platform for achieving this requirement is a stabilizer code[42], in which the wavefunctions are characterized by a set of commuting operators $\{{\hat{S}}_{p}\}$ called stabilizers, with ${\hat{S}}_{p}\left|\psi \right\rangle ={s}_{p}\left|\psi \right\rangle$ and $s_p = \pm 1$. The code space is the set of degenerate wavefunctions for which $s_p = 1$ for all p. Hence, every independent stabilizer divides the degeneracy of the code space by two.

Whereas the physical layout of qubits is typically used to determine the structure of the stabilizers, the qubits can be considered to be degree j vertices (DjV; $j \in \{2, 3, 4\}$) on more general planar graphs (Fig. 1a)[23]. Using this picture, each stabilizer can be associated with a plaquette p, whose vertices are the qubits on which ${\hat{S}}_{p}$ acts:

$${\hat{S}}_{p}=\prod _{v\,\in \,{\rm{vertices}}}{\hat{\tau }}_{p,v}.$$
(1)

${\hat{\tau }}_{p,v}$ is here a single-qubit Pauli operator acting on vertex *v*, chosen to satisfy a constraint around that vertex (Fig. 1b). An instance where $s_p = -1$ on a plaquette is called a plaquette violation. These can be thought of as quasiparticles, which are created and moved through single-qubit Pauli operators (Fig. 1a). A pair of plaquette violations sharing an edge constitute a fermion, ε. We recently demonstrated the Abelian statistics of such quasiparticles in the surface code[43]. To realize non-Abelian statistics, one needs to go beyond such plaquette violations; it has been proposed that dislocations in the stabilizer graph—analogous to lattice defects in crystalline solids—can host projective non-Abelian Ising anyons[9,10]. For brevity, we refer to these as ‘non-Abelian anyons’ or simply ‘anyons’ from here on.

Fig. 1: Deformations of the surface code.



a, Stabilizer codes are conveniently described in a graph framework. Through deformations of the surface code graph, a square grid of qubits (crosses) can be used to realize more generalized graphs. Plaquette violations (red) correspond to stabilizers with $s_p = -1$ and are created by local Pauli operations. In the absence of deformations, plaquette violations are constrained to move on one of the two sublattices of the dual graph in the surface code, hence the two shades of blue. **b**, A pair of D3Vs (yellow triangles) appears by removing an edge between two neighbouring

stabilizers, ${\hat{S}}_{1}$ and ${\hat{S}}_{2}$, and introducing the new stabilizer, $\hat{S}={\hat{S}}_{1}{\hat{S}}_{2}$. A D3V is moved by applying a two-qubit entangling gate, $\exp \left(\frac{\pi }{8}[{\hat{S}}^{{\prime} },\hat{S}]\right)$. In the presence of bulk D3Vs, there is no consistent way of chequerboard colouring, hence the (arbitrarily chosen) grey regions. The top right shows that in a general stabilizer graph, ${\hat{S}}_{p}$ can be found from a constraint at each vertex, where $\{\tau_1, \tau_2\} = 0$.

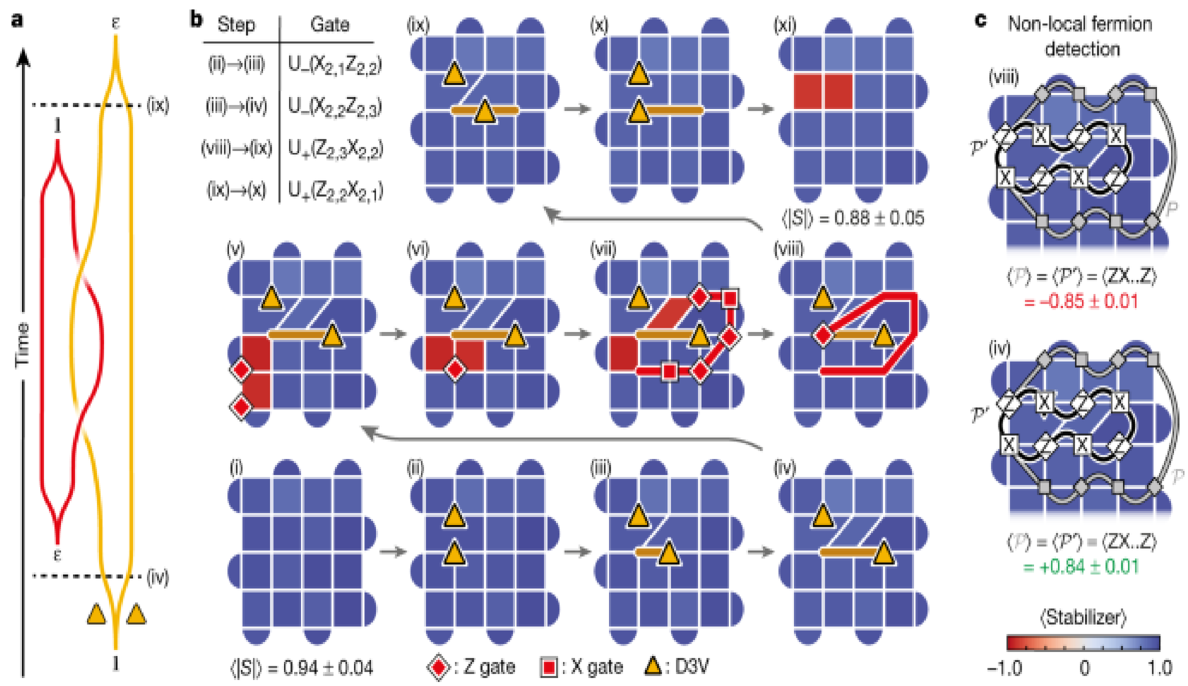
In the graph framework introduced above, it has been shown that such dislocations are characterized as vertices of degree 2 and 3 (ref. [23]). Consider the stabilizer graph of the surface code[26,44], specifically with boundary conditions such that the degeneracy is two. Although all the vertices in the bulk are D4Vs, one can create two D3Vs by removing an edge between two neighbouring plaquettes p and q, and introducing the new stabilizer $\hat{S}={\hat{S}}_{p}{\hat{S}}_{q}$ (Fig. 1b). Evidently, the introduction of two D3Vs reduces the number of independent stabilizers by one and thus doubles the degeneracy. This doubling is exactly what is expected when a pair of Ising anyons is introduced[9,10]; hence, D3Vs appear as a candidate of non-Abelian anyons, and we will denote them as σ.

To be braided and fused by unitary operations, the D3Vs must be moved. Whereas the structure of the stabilizer graph is usually considered to be static, it was predicted by Bombin that the dislocations in the surface code would show projective non-Abelian Ising statistics if braided[10]. Here, we will use a specific protocol recently proposed by Lensky et al.[23] for deforming the stabilizer graph (and thus moving the anyons) using local two-qubit Clifford gates. To shift a D3V from vertex u to v, an edge must be disconnected from v and reconnected to u. This can be achieved by means of the gate unitary $\exp \left(\frac{\pi }{8}[{\hat{S}}_{p}^{{\prime} },{\hat{S}}_{p}]\right)$, where ${\hat{S}}_{p}$ is the original stabilizer containing the edge and u, and ${\hat{S}}_{p}^{{\prime} }$ is the new stabilizer that emerges after moving the edge[23]. In cases where the D3V is shifted between two connected vertices, the unitary simplifies to the form ${U}_{\pm }({\hat{\tau }}_{u}{\hat{\tau }}_{v})\equiv \exp \left(\pm i\frac{\pi }{4}{\hat{\tau }}_{u}{\hat{\tau }}_{v}\right)$, where ${\hat{\tau }}_{u}$ and ${\hat{\tau }}_{v}$ are Pauli operators acting on

vertices u and v. We experimentally realize this unitary through a controlled Z (CZ) gate and single-qubit rotations (median errors of 7.3×10^{-3} and 1.3×10^{-3}, respectively; Methods).

Following these insights from Kitaev and Bombin, we now turn to our experimental study of the proposed anyons, using the protocol described in ref. [23]. In the first experiment, we demonstrate the creation of anyons and the fundamental fusion rules of σ and ε (Fig. 2a). In a 5 × 5 grid of superconducting qubits, we first use a protocol consisting of four layers of CZ gates to prepare the surface code ground state (Fig. 2b(i), see also ref. [43]). The average stabilizer value after the ground state preparation is 0.94 ± 0.04 (individual stabilizer values shown in Extended Data Fig. 3c). We then remove a stabilizer edge to create a pair of D3Vs (σ) and separate them through the application of two-qubit gates. Fig. 2b(i)–(iv) show the measured stabilizer values in the resultant graph in each step of this procedure (determined by simultaneously measuring the involved qubits in their respective bases, n = 10,000; note that the measurements are destructive and the protocol is restarted after each measurement). In Fig. 2b(v), single-qubit Z gates are applied to two qubits near the lower left corner of the grid to create adjacent plaquette violations, which together form a fermion. Through the sequential application of X and Z gates (Fig. 2b(vii)–(viii)), one of the plaquette violations is then made to encircle the right σ vertex. Crucially, after moving around σ, the plaquette violation does not return to where it started, but rather to the location of the other plaquette violation. This enables them to annihilate (Fig. 2b(viii)), causing the fermion to seemingly disappear. However, by bringing the two σ back together and annihilating them (Fig. 2b(ix)–(xi)), we arrive at a striking observation: an ε particle re-emerges on two of the square plaquettes where the σ vertices previously resided.

Fig. 2: Demonstration of the fundamental fusion rules of D3Vs.



a, The braiding worldlines used to fuse ε and σ. **b**, Expectation values of stabilizers at each step of the unitary operation after readout correction (see Extended Data Fig. 3 for details and individual stabilizer values). We first prepare the ground state of the surface code (step (i); average stabilizer value of 0.94 ± 0.04, where the uncertainty is one standard deviation). A D3V (σ) pair is then created (ii) and separated (iii)–(iv), before creating a fermion, ε (v). One of the plaquette violations is brought around the right σ (vi)–(viii), allowing it to annihilate with the other plaquette violation (viii). The fermion has seemingly disappeared, but re-emerges when the σ are annihilated ((xi); stabilizer values −0.86 and −0.87). The path (v) → (viii) demonstrates the fusion rule, $\sigma \times \varepsilon = \sigma$. The different fermion parities at the end of the paths (viii) → (xi) and (iv) → (i) show the other fusion rule, $\sigma \times \sigma ={\mathbb{1}}+\varepsilon$. Yellow triangles represent the positions of the σ. The brown and red lines denote the paths of the σ and the plaquette violation, respectively. Red squares (diamonds) represent X (Z) gates. Upper left shows a table of two-qubit unitaries used in the protocol. Each stabilizer was measured n = 10,000 times in each step. **c**, A non-local technique for hidden fermion detection: the presence of a fermion in a σ-pair can be deduced by measuring the sign of the Pauli string $\hat{{\mathcal{P}}}$ corresponding to bringing a plaquette violation around the σ-pair (grey path). $\hat{{\mathcal{P}}}$ is equivalent to the shorter string $\hat{{{\mathcal{P}}}^{{\prime} }}$ (black path). Measurements of \

$\hat{{{\mathcal{P}}}^{{\prime} }}$ in steps (viii) (top) and (iv) (bottom) give values of −0.85 ± 0.01 and +0.84 ± 0.01, respectively. This indicates that there is a hidden fermion pair in the former case, but not in the latter, despite the stabilizers being the same.

Our results demonstrate the fusion of ε and σ. The disappearance of the fermion from step (v) to (viii) establishes the fundamental fusion rule of ε and σ:

$$\sigma \times \varepsilon =\sigma .$$
(2)

We emphasize that none of the single-qubit gates along the path of the plaquette violation are applied to the qubits hosting the mobile σ; our observations are therefore solely due to the non-local effects of non-Abelian D3Vs, and exemplify the unconventional behaviour of the latter. Moreover, another fusion rule is seen by considering the reverse path (iv) → (i), and comparing it to the path (viii) → (xi). These two paths demonstrate that a pair of σ can fuse to form either vacuum (${\mathbb{1}}$) or one fermion (steps (i) and (xi), respectively):

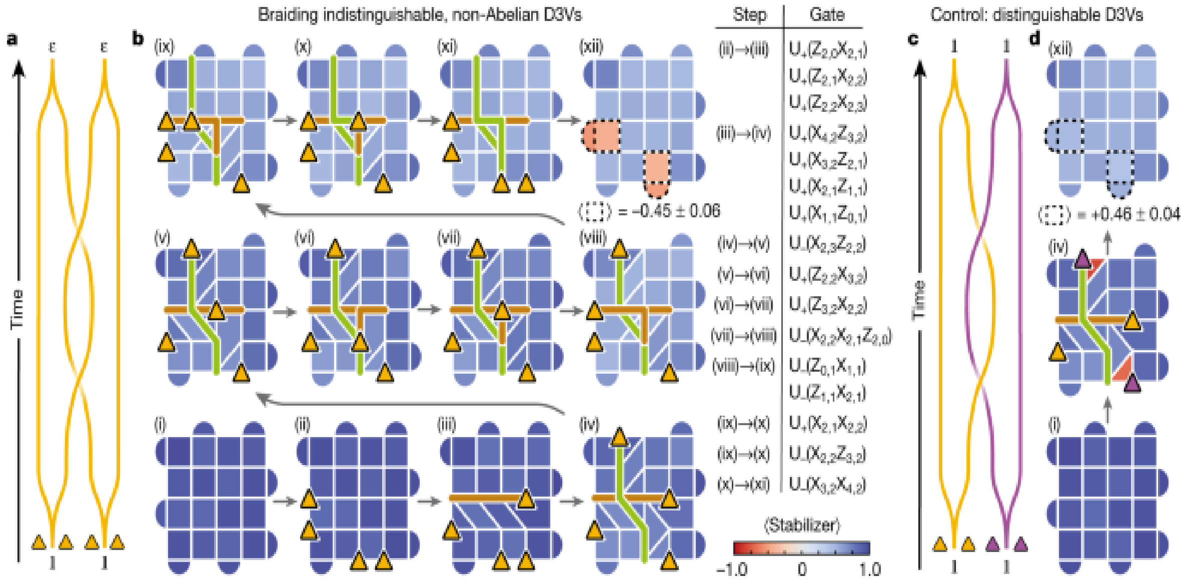
$$\sigma \times \sigma ={\mathbb{1}}+\varepsilon .$$
(3)

The starting points of these two paths ((iv) and (viii)) cannot be distinguished by any local measurement. We therefore introduce a non-local measurement technique that allows for detecting an ε without fusing the σ (refs. [10,23,26]). The key idea underlying this method is that bringing a plaquette violation around a fermion should result in a π phase. We therefore measure the Pauli string $\hat{{\mathcal{P}}}$ that corresponds to creating two plaquette violations, bringing one of them around the two σ, and finally annihilating them with each other (grey paths in Fig. 2c). The existence of an ε inside the σ-pair should cause $\langle \hat{{\mathcal{P}}}\rangle =-1$. To simplify this technique further, $\hat{{\mathcal{P}}}$ can be reduced to a shorter string $\hat{{{\mathcal{P}}}^{{\prime} }}$ (black paths in Fig. 2c) by taking advantage of the stabilizers it encompasses. For instance, if $\hat{{\mathcal{P}}}$ contains three of the operators in a four-qubit stabilizer, these can be switched out with the remaining operator.

Measuring $\langle \hat{{{\mathcal{P}}}^{{\prime} }}\rangle$ in step (iv), in which the σ are separated but the fermion has not yet been introduced, gives $\langle \hat{{{\mathcal{P}}}^{{\prime} }}\rangle =+\,0.84\pm 0.01$, consistent with the absence of fermions (Fig. 2c). However, performing the exact same measurement in step (viii), in which the σ are in the same positions, we find $\langle \hat{{{\mathcal{P}}}^{{\prime} }}\rangle =-\,0.85\pm 0.01$, indicating that an ε is delocalized across the spatially separated σ pair (Fig. 2c). This observation highlights the non-local encoding of the fermions, which cannot be explained with classical physics.

Having demonstrated the above fusion rules involving σ, we next braid them with each other to directly show their non-Abelian statistics. We consider two spatially separated σ pairs, A and B, by removing two stabilizer edges (Fig. 3a,b(ii)). Next, we apply two-qubit gates along a horizontal path to separate the σ in pair A (Fig. 3b(iii)), followed by a similar procedure in the vertical direction on pair B (Fig. 3b(iv)), so that one of its σ crosses the path of pair A. We then subsequently bring the σ from pairs A and B back to their original positions (Fig. 3b(v)–(viii) and (ix)–(xi), respectively). When the two σ pairs are annihilated in the final step (Fig. 3b(xii)), we observe that a fermion is revealed in each of the positions where the σ pairs resided (average stabilizer value -0.45 ± 0.06). This shows a clear change in local observables from the initial state in which no fermions were present. As a control experiment, we repeat the experiment with distinguishable σ pairs, achieved by attaching a plaquette violation to each of the σ in pair B (Fig. 3c,d; see also Extended Data Fig. 8 for stabilizer measurements through the full protocol). Moving the plaquette violation along with the σ requires a string of single-qubit gates, which switches the direction of the rotation in the multiqubit unitaries, $U_{\pm} \to U_{\mp}$. In this case, no fermions are observed at the end of the protocol (average stabilizer value $+0.46 \pm 0.04$), thus providing a successful control.

Fig. 3: Braiding of non-Abelian anyons.



a, Wordline schematic of the braiding process. **b**, Experimental demonstration of braiding, showing the values of the stabilizers throughout the process. Two σ pairs, A and B, are created from the vacuum \({\mathbb{1}}\), and one of the σ in pair A is brought to the right side of the grid. Next, a σ from pair B is moved to the top, thus crossing the path of pair A, before bringing σ pairs A and B back together to complete the braid. In the final step, two fermions appear in the locations where the σ pairs resided, constituting a change in the local observables. The diagonal σ move in step (iv) requires two SWAP gates (three CZ gates each) and a total of ten CZ gates. The three-qubit unitary in step (viii) requires four SWAP gates and a total of 15 CZ gates. In the full circuit, a total of 40 layers of CZ gates are applied (Methods). The yellow triangles represent the locations of the σ; the brown and green lines represent the paths of σ from pairs A and B, respectively. The four red stabilizers in (xii) have a mean value of -0.45 ± 0.06, where the uncertainty is one standard deviation. Each stabilizer was measured $n = 10,000$ times in each step. **c**, As a control experiment, we perform the same braid as in **a**, but with distinguishable σ by attaching a plaquette violation to the σ in pair B (represented with purple triangles). **d**, Same as **b**, but using distinguishable σ (only showing steps (i), (iv) and (xii)). In contrast to **b**, no fermions are observed in step (xii).

Fermions can only be created in pairs in the bulk. Moreover, the fusion of two σ can only create zero or one fermion (equation (3)). Hence, our experiment involves the minimal number of bulk σ (four) needed to encode

two fermions and demonstrate non-Abelian braiding. Because the fermion parity is conserved, effects of gate imperfections and decoherence can be partially mitigated by postselecting for an even number of fermions. This results in fermion detection values of −0.76 ± 0.03 and +0.79 ± 0.04 in Fig. 3b,d, respectively.

Together, our observations show the change in local observables by braiding of indistinguishable σ and constitute a direct demonstration of their non-Abelian statistics. In other words, the double-braiding operation R^2 is a matrix that cannot be reduced to a scalar. Specifically, it corresponds to an X gate acting on the space spanned by zero- and two-fermion wavefunctions.

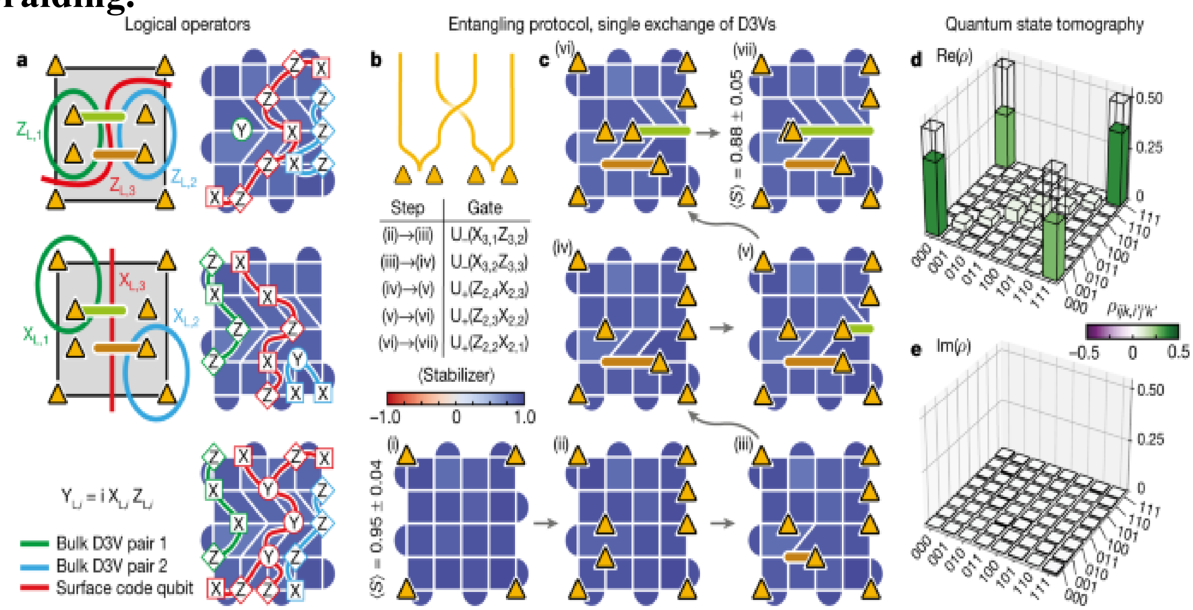
The full braiding circuit consists of 40 layers of CZ gates and 41 layers of single-qubit gates (36 of each after ground state preparation). The effects of imperfections in this hardware implementation can be assessed through comparison with the control experiment. The absolute values of the stabilizers in which the fermions are detected in the two experiments (dashed boxes in Fig. 3b,d(xii)) are very similar (average values of −0.45 and +0.46). This is consistent with the depolarization channel model, in which the measured stabilizer values are proportional to the ideal values (±1).

We next study the prospects of using D3Vs to encode logical qubits and prepare an entangled state of anyon pairs. By doubling the degeneracy, each additional σ pair introduces one logical qubit, where the ${\left|0\right\rangle }_{{\rm{L}}}$ (${| 1\rangle }_{{\rm{L}}}$) state corresponds to an even (odd) number of hidden fermions. The measurements of the fermion numbers in several σ pairs are not fully independent: bringing a plaquette violation around one σ pair is equivalent to bringing it around all the other pairs (due to the conservation of fermionic parity). Hence, $N \geq 2$ anyons encode $N/2 - 1$ logical qubits. The D3Vs we have created and manipulated so far are not the only ones present in the stabilizer graph; with the boundary conditions used here, each of the four corners are also D3Vs, no different from those in the bulk[23]. Indeed, the existence of D3Vs in the corners is the reason why a single fermion could be created in the corner in Fig. 2b(v). This is also consistent with the fact that the surface code itself encodes one logical qubit in the absence of additional D3Vs. Here we create

two pairs of D3Vs, in addition to the four that are already present in the corners, to encode a total of three logical qubits.

Through the use of braiding, we aim to prepare an entangled state of these logical qubits, specifically a GHZ state on the form $(| 000\rangle +| 111\rangle)/\sqrt{2}$. The definition of a GHZ state and the specifics of how it is prepared is basis-dependent. In most systems, the degrees of freedom are local and there is a natural choice of basis. For spatially separated anyons, the measurement operators are necessarily non-local. Here we choose the basis defined as follows: for the first two logical qubits, we choose the logical ${\hat{Z}}_{{\rm{L}},i}$ operators to be Pauli strings encircling each of the bulk σ pairs, as was used in Fig. 2c (green and turquoise paths in the left column of Fig. 4a). For the logical surface code qubit, we define ${\hat{Z}}_{{\rm{L}},3}$ as the Pauli string that crosses the grid horizontally through the gap between the bulk D3V pairs, effectively enclosing four σ (red path in Fig. 4a). In this basis, the initial state is a product state.

Fig. 4: Entangled state of anyon-encoded logical qubits by means of braiding.



a, Logical operators of the three logical qubits encoded in the eight anyons (other basis choices are possible). The coloured curves in the left column denote plaquette violation paths, before reduction to shorter, equivalent Pauli strings measured in the experiment (right column). **b**, Worldline schematic

of the single exchange used to realize an entangled state of the logical qubits. **c**, Single exchange of the non-Abelian anyons, showing measurements of the stabilizers throughout the protocol. Yellow triangles represent the locations of the *σ*, whereas brown and green lines denote their paths. The average stabilizer values are 0.95 ± 0.04 and 0.88 ± 0.05 (one standard deviation) in the first and last step, respectively. Each stabilizer was measured n = 20,000 times in each step. **d**,**e**, Real (**d**) and imaginary (**e**) parts of the reconstructed density matrix from the quantum state tomography. ${\rm{Re}}(\rho)$ has clear peaks in its corners, as expected for a GHZ state on the form $(\left|000\right\rangle +\left|111\right\rangle)/\sqrt{2}$. The overlap with the ideal GHZ state is ${\rm{Tr}}\{\,{\rho }_{{\rm{GHZ}}}\rho \}=0.623\pm 0.004$, where the uncertainty is one standard deviation determined from bootstrapping.

Whereas a double braid was used to implement the operator X in Fig. 3, we now perform a single braid (Fig. 4b) to realize $\sqrt{X}$ and create a GHZ state. We implement this protocol by bringing one *σ* from each bulk pair across the grid to the other side (Fig. 4c). For every anyon double exchange across a Pauli string, the value of the Pauli string changes sign. Hence, a double exchange would change $\left|000\right\rangle$ to $\left|111\right\rangle$, whereas a single exchange is expected to realize the superposition, $(\left|111\right\rangle +\left|000\right\rangle)/\sqrt{2}$.

To study the effect of this operation, we perform quantum state tomography on the final state, which requires measurements of not only ${\hat{Z}}_{{\rm{L}},i}$, but also ${\hat{X}}_{{\rm{L}},i}$ and ${\hat{Y}}_{{\rm{L}},i}$ on the three logical qubits. For the first two logical qubits, ${\hat{X}}_{{\rm{L}},i}$ is the Pauli string that corresponds to bringing a plaquette violation around only one of the *σ* in the pair (as demonstrated in Fig. 2b). Both the logical ${\hat{X}}_{{\rm{L}},i}$ and ${\hat{Z}}_{{\rm{L}},i}$ operators can be simplified by reducing the original Pauli strings (green and turquoise paths in the left column of Fig. 4c) to equivalent, shorter ones (right column). ${\hat{Z}}_{{\rm{L}},1}$ can in fact be reduced to a single $\hat{Y}$-operator. For the logical surface code qubit, we define ${\hat{X}}_{{\rm{L}},3}$ as the Pauli string that crosses the grid vertically between the bulk D3V pairs (red path in Fig. 4a). Finally, the logical ${\hat{Y}}_{{\rm{L}},i}$-operators are simply found from

$({\hat{Y}}_{{\rm{L}},i}=i{\hat{X}}_{{\rm{L}},i}{\hat{Z}}_{{\rm{L}},i})$. Measuring these operators, we reconstruct the density matrix of the final state (Fig. 4d,e), which has a purity of $\sqrt{{\rm{Tr}}\{{\rho }^{2}\}}=0.646\pm 0.003$ and an overlap with the ideal GHZ state of ${\rm{Tr}}\{{\rho }_{{\rm{GHZ}}}\rho \}=0.623\pm 0.004$ (uncertainties estimated from bootstrapping method; resampled 10,000 times from the original data set). The fact that the state fidelity is similar to the purity suggests that the infidelity is well described by a depolarizing error channel.

In conclusion, we have realized highly controllable braiding of degree-3 vertices, enabling the demonstration of the fusion and braiding rules of non-Abelian Ising anyons. We have also shown that braiding can be used to create an entangled state of three logical qubits encoded in these anyons. In other, more conventional candidate platforms for non-Abelian exchange statistics, which involve Hamiltonian dynamics of quasi-particle excitations, topological protection naturally arises from an emergent gap that separates the computational states from other states. To leverage the non-Abelian anyons in our system for topologically protected quantum computing, the stabilizers must be measured throughout the braiding protocol. The potential inclusion of this error correction procedure, which involves overheads including readout of five-qubit stabilizers, could open a new path towards fault-tolerant implementation of Clifford gates, a key ingredient of universal quantum computation.

Methods

Qubit decoherence and gate characterization

The experiments are performed on a quantum processor with frequency-tuneable transmon qubits and a similar design to that in ref. [45]. Extended Data Fig. 1a shows the measured relaxation times of the 25 qubits that were used in the experiment, with a median value of $T_1 = 21.7$ µs. We also measure the dephasing time T_2 in a Hahn echo experiment, shown in Extended Data Fig. 1b, with the same median value of 21.7 µs. We note that the equality of T1 and T2 is a coincidence and that the discrepancy between

the measured decoherence rate $1/T_2$ and the relaxation-limited rate $1/(2T_1)$ is due to remnant noise not decoupled in the Hahn echo experiment.

Next, we benchmark the gates used in the experiment. Extended Data Fig. 2a,b show the cumulative distribution of the Pauli errors for single- and two-qubit (CZ) gates, respectively. The median Pauli errors are 1.3×10^{-3} for the single-qubit gates and 7.3×10^{-3} for the two-qubit gates.

Readout details

Because the readout of the qubit state is imperfect, the raw data gives a somewhat incorrect representation of the actual quantum state of the system. We write the probability of readout error of state 0(1) on qubit i as $p_{0(1),i}$, and the readout fidelity is thus given by $1-(p_{0,i}+p_{1,i})/2$. To correct for any asymmetry between readout of the $\left|0\right\rangle$ and $\left|1\right\rangle$ states, we perform symmetrized measurements in which π pulses are applied to the qubits before the readout in half of the measurements and the recorded qubit values are inverted. The measured value of a stabilizer with actual value $\langle S\rangle = \langle \prod_i \alpha_i \rangle$ (where the product runs over qubits in the stabilizer) is then:

$${\langle S\rangle }_{{\rm{meas}}}=\langle \prod _{i}(1-{p}_{0,i}-{p}_{1,i}){\alpha }_{i}\rangle =\prod _{i}(1-{p}_{0,i}-{p}_{1,i})\langle S\rangle ,$$
(4)

where we made use of the fact that each qubit is measured equally often in the $\left|0\right\rangle$ and $\left|1\right\rangle$ states in the symmetrized measurements. Note the absence of the factor 1/2 compared to the expression for the readout fidelity, as perfectly incorrect readout ($p_0 = p_1 = 1$) would give a readout fidelity of 0, but a measured value of $-\alpha_i$. To correct for the discrepancy between the measured stabilizer value and the actual stabilizer value, we measure $\langle Z_1 ... Z_n \rangle$ of the state $\left|00..00\right\rangle$ with the same qubits (using again symmetrized measurements) to find:

$${\langle {Z}_{1}..{Z}_{n}\rangle }_{{\rm{meas}}}=\prod _{i}1-{p}_{0,i}-{p}_{1,i}$$
(5)

The readout-corrected ⟨S⟩ is then found from:

$${\langle S\rangle }_{{\rm{corr}}}={\langle S\rangle }_{{\rm{meas}}}/{\langle {Z}_{1}..{Z}_{n}\rangle }_{{\rm{meas}}}$$
(6)

Extended Data Fig. 3 shows the measured readout errors, as well as a comparison of the stabilizer values in the surface code ground state (same data as Fig. 2b(i)) before and after readout correction.

Dynamical decoupling

To mitigate the effects of qubit decoherence during the circuits, we perform dynamical decoupling on qubits that are idle for more than three layers of gates. In particular, we use the Carr–Purcell–Meiboom–Gill scheme, consisting of X pulses interspaced by a wait time of $\tau = 25$ ns. Extended Data Fig. 4 shows an example comparison of the stabilizers in cases with and without dynamical decoupling, after braiding of anyons (41 layers of SQ gates and 40 layers of CZ gates). A clear improvement is observed, increasing the average absolute stabilizer value from 0.50 to 0.58.

Circuit details

Extended Data Fig. 5 shows the circuits used in the experiments presented in the main text. In our experiment, the two-qubit unitaries $({U}_{\pm }({\hat{\tau }}_{1}{\hat{\tau }}_{2}))$ are converted to single-qubit rotations and CZ gates, as shown in Extended Data Fig. 6b. In the particular case in which a D3V is moved diagonally (Fig. 3b(iv)), we realize the unitary by including two SWAP gates (also converted to CZ gates) as the qubits are connected in a square grid (Extended Data Fig. 6c). Moreover, the three-qubit unitary in Fig. 3b(viii) is equivalent to a combination of single-qubit gates, four SWAP gates and four CZ gates (Extended Data Fig. 6d), which can be further converted to single-qubit gates and 15 CZ gates. In the experimental implementation of the circuit, adjacent single-qubit gates on

the same qubit are merged and performed in the layer after the most recent CZ gate (Extended Data Fig. 6e).

Numerical simulation of braiding in presence of noise

To better understand the role of errors in the experimental results in Fig. 3 of the main text, we perform a numerical simulation of the density matrix evolution subject to the braiding circuit in the presence of noise. We use the method of quantum trajectories to approximate the expectation value of stabilizers with the 25-qubit density matrix. The model of noise includes T_1 and T_2 effects described by the single-qubit Kraus operators,

$${K}_{0}=\left(\begin{array}{cc}1 & 0\\ 0 & \exp (-t/{T}_{2})\end{array}\right),$$
(7)

$${K}_{1}=\left(\begin{array}{cc}0 & \sqrt{1-\exp (-t/{T}_{1})}\\ 0 & 0\end{array}\right),$$
(8)

$${K}_{2}=\left(\begin{array}{cc}0 & 0\\ 0 & \sqrt{\exp (-t/{T}_{1})-\exp (-2t/{T}_{2})}\end{array}\right),$$
(9)

where t is the duration of the evolution, as well as additional one- and two-qubit depolarizing channel error for each gate. The depolarizing channel error rate is chosen such that the combined Pauli error from T_1, T_2 and depolarizing error matches the gate Pauli error measured in an independent experiment (Qubit decoherence and gate characterization). We take these values to be uniform across the chip. The expectation values of the four stabilizers that correspond to the noise-free value of −1, see light red stabilizers in Fig. 3b(xii) and Extended Data Fig. 7, have the following values (×100): (−58, −46, −34, −46) with statistical error 4. For comparison, the experimental values for the same set of stabilizers is (−52, −41, −39, −49). Our simulation results are in relatively good agreement with the measured data, suggesting that the model captures the effects of noise well. The observed discrepancies are expected to be due to inhomogeneity of the errors, which was not included in our error model. The simulations used an open source simulator qsim[46].

Additional braiding data

In Fig. 3, we demonstrate that no fermion appears when distinguishable σ are braided with each other. In Extended Data Fig. 8, we show the data for each step in that protocol, analogous to those shown for indistinguishable σ in the main text. Moreover, we also present an alternative braiding scheme in Extended Data Fig. 9, which requires fewer (18) CZ gates. In this case, however, pair B is not brought back together, and neither of the σ pairs are annihilated. Therefore, similar to in Fig. 2c, we measure the Pauli string corresponding to bringing a plaquette violation around pair A (grey path in Extended Data Fig. 9c), which in this case can be reduced to $\hat{Y}$ on the qubit where the two σ overlap. We find $\langle \hat{{\mathcal{P}}}\rangle =\langle \hat{Y}\rangle =-\,0.71\,\pm \,0.01$, indicating that braiding the σ led to the creation of a fermion (Extended Data Fig. 9c). Note that we here only search for fermions in one of the σ pairs. As a control experiment, we repeat the experiment with distinguishable σ pairs, as in the main text (Extended Data Fig. 9d). In this case, we find $\langle \hat{{\mathcal{P}}}\rangle =+\,0.71\,\pm \,0.01$, thus demonstrating that no fermion was produced. Together, these observations constitute another demonstration of non-Abelian exchange statistics of the D3Vs.

A summary of the theoretical framework

It was observed by Kitaev that fluxes of the e–m exchange symmetry are expected to host Majorana modes and therefore have the degeneracy of Ising anyons[9]. Bombin gave a particular stabilizer configuration realizing such a flux as a fixed lattice dislocation in a square grid, showed on general grounds that if such fluxes were well-separated and could be braided they would be projective Ising anyons, and noted that it may be possible to braid such fluxes by code deformation[10]. A general formalism for theories realized by braiding of symmetry fluxes was described in ref. [47]. These constructions focus on the long-distance physics, and in practical terms[23] gives an account of ‘microscopics’. An explicit mapping to a gauge theory shows how the anyons are localized to a single qubit, and is used to derive a simple, efficient and systematic procedure for creating, braiding and measuring the fusion outcomes of Ising anyons on general stabilizer graphs. The bridge between the microscopics and general arguments established by the gauge

theory mapping allows us to fit several anyons on present-day devices, probe the full two-dimensional nature of their braiding by maintaining their separation, and demonstrate braid generators that restore all local observables. For details discussions of the protocol, see ref. [23].

Data availability

The data that support the findings in this study are available at https://doi.org/10.5281/zenodo.7869220.

References

1. Sakurai, J. J. *Modern Quantum Mechanics* (Addison-Wesley, 1993).

2. Leinaas, J. & Myrheim, J. On the theory of identical particles. *Nuovo Cim. B.* **37**, 1–23 (1977).

3. Wilczek, F. Quantum mechanics of fractional-spin particles. *Phys. Rev. Lett.* **49**, 957–959 (1982).

4. Wilczek, F. *Fractional Statistics and Anyon Superconductivity* (World Scientific, 1990).

5. Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Sarma, S. D. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80**, 1083–1159 (2008).

6. Camino, F. E., Zhou, W. & Goldman, V. J. $e/3$ Laughlin quasiparticle primary-filling $\nu = 1/3$ interferometer. *Phys. Rev. Lett.* **98**, 076805 (2007).

7. Bartolomei, H. et al. Fractional statistics in anyon collisions. *Science* **368**, 173–177 (2020).

8. Nakamura, J., Liang, S., Gardner, G. C. & Manfra, M. J. Direct observation of anyonic braiding statistics. *Nat. Phys.* **16**, 931—936 (2020).

9. Kitaev, A. Anyons in an exactly solved model and beyond. *Ann. Phys.* **321**, 2–111 (2006).

10. Bombin, H. Topological order with a twist: Ising anyons from an Abelian model. *Phys. Rev. Lett.* **105**, 030403 (2010).

11. Hormozi, L., Zikos, G., Bonesteel, N. E. & Simon, S. H. Topological quantum compiling. *Phys. Rev. B* **75**, 165310 (2007).

12. You, Y.-Z. & Wen, X.-G. Projective non-abelian statistics of dislocation defects in a \({{\mathbb{Z}}}_{N}\) rotor model. *Phys. Rev. B* **86**, 161107 (2012).

13. Barkeshli, M., Jian, C.-M. & Qi, X.-L. Twist defects and projective non-Abelian braiding statistics. *Phys. Rev. B* **87**, 045130 (2013).

14. Barkeshli, M. & Qi, X.-L. Topological nematic states and non-Abelian lattice dislocations. *Phys. Rev. X* **2**, 031013 (2012).

15. von Keyserlingk, C. W., Burnell, F. J. & Simon, S. H. Three-dimensional topological lattice models with surface anyons. *Phys. Rev. B* **87**, 045107 (2013).

16. Teo, J. C., Roy, A. & Chen, X. Unconventional fusion and braiding of topological defects in a lattice model. *Phys. Rev. B* **90**, 115118 (2014).

17. Zheng, H., Dua, A. & Jiang, L. Demonstrating non-Abelian statistics of majorana fermions using twist defects. *Phys. Rev. B* **92**, 245139 (2015).

18. Teo, J. C., Hughes, T. L. & Fradkin, E. Theory of twist liquids: gauging an anyonic symmetry. *Ann. Phys.* **360**, 349–445 (2015).

19. Brown, B. J., Laubscher, K., Kesselring, M. S. & Wootton, J. R. Poking holes and cutting corners to achieve clifford gates with the surface code. *Phys. Rev. X* **7**, 021029 (2017).

20. Zhu, G., Hafezi, M. & Barkeshli, M. Quantum origami: transversal gates for quantum computation and measurement of topological order. *Phys. Rev. Res.* **2**, 013285 (2020).

21. Benhemou, A., Pachos, J. K. & Browne, D. E. Non-abelian statistics with mixed-boundary punctures on the toric code. *Phys. Rev. A* **105**, 042417 (2022).

22. Tantivasadakarn, N., Verresen, R. & Vishwanath, A. The shortest route to non-Abelian topological order on a quantum processor. Preprint at https://arxiv.org/abs/2209.03964 (2022).

23. Lensky, Y. D., Kechedzhi, K., Aleiner, I. & Kim, E.-A. Graph gauge theory of mobile non-Abelian anyons in a qubit stabilizer code. Preprint at https://arxiv.org/abs/2210.09282 (2022).

24. Stern, A. Anyons and the quantum Hall effect: a pedagogical review. *Ann. Phys*. **323**, 204–249 (2008).

25. Harle, N., Shtanko, O. & Movassagh, R. Observing and braiding topological Majorana modes on programmable quantum simulators. *Nat. Commun.* **14**, 2286 (2023)

26. Kitaev, A. Y. Fault-tolerant quantum computation by anyons. *Ann. Phys*. **303**, 2–30 (2003).

27. Freedman, M. H. P/np, and the quantum field computer. *Proc. Natl Acad. Sci. USA* **95**, 98–101 (1998).

28. Pachos, J. K. *Introduction to Topological Quantum Computation* (Cambridge Univ. Press, 2012).

29. Stern, A. & Lindner, N. Topological quantum computation—from basic concepts to first experiments. *Science* **339**, 1179–1184 (2013).

30. Field, B. & Simula, T. Introduction to topological quantum computation with non-Abelian anyons. *Quantum Sci. Technol.* **3**, 045004 (2018).

31. Moore, G. & Read, N. Nonabelions in the fractional quantum Hall effect. *Nucl. Phys. B* **360**, 362–396 (1991).

32. Willett, R. L., Pfeiffer, L. N. & West, K. W. Alternation and interchange of e/4 and e/2 period interference oscillations consistent with filling factor 5/2 non-Abelian quasiparticles. *Phys. Rev. B* **82**, 205301 (2010).

33. Read, N. & Green, D. Paired states of fermions in two dimensions with breaking of parity and time-reversal symmetries and the fractional quantum Hall effect. *Phys. Rev. B* **61**, 10267 (2000).

34. Ivanov, D. A. Non-Abelian statistics of half-quantum vortices in *P*-wave superconductors. *Phys. Rev. Lett.* **86**, 268–271 (2001).

35. Lutchyn, R. M., Sau, J. D. & Das Sarma, S. Majorana fermions and a topological phase transition in semiconductor-superconductor heterostructures. *Phys. Rev. Lett.* **105**, 077001 (2010).

36. Oreg, Y., Refael, G. & von Oppen, F. Helical liquids and majorana bound states in quantum wires. *Phys. Rev. Lett.* **105**, 177002 (2010).

37. Mourik, V. et al. Signatures of majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* **336**, 1003–1007 (2012).

38. Nadj-Perge, S. et al. Observation of majorana fermions in ferromagnetic atomic chains on a superconductor. *Science* **346**, 602–607 (2014).

39. Banerjee, M. et al. Observation of half-integer thermal hall conductance. *Nature* **559**, 205–210 (2018).

40. Kasahara, Y. et al. Majorana quantization and half-integer thermal quantum Hall effect in a Kitaev spin liquid. *Nature* **559**, 227–231 (2018).

41. Bonderson, P., Kitaev, A. & Shtengel, K. Detecting non-Abelian statistics in the $\nu = 5/2$ fractional quantum Hall state. *Phys. Rev. Lett.* **96**, 016803 (2006).

42. Gottesman, D. *Stabilizer Codes and Quantum Error Correction* (California Institute of Technology, 1997).

43. Satzinger, K. et al. Realizing topologically ordered states on a quantum processor. *Science* **374**, 1237–1241 (2021).

44. Wen, X.-G. Quantum orders in an exact soluble model. *Phys. Rev. Lett.* **90**, 016803 (2003).

45. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).

46. Isakov, S. V. et al. Simulations of quantum circuits with approximate noise using qsim and cirq. Preprint at https://arxiv.org/abs/2111.02396 (2021).

47. Barkeshli, M., Bonderson, P., Cheng, M. & Wang, Z. Symmetry fractionalization, defects, and gauging of topological phases. *Phys. Rev. B* **100**, 115147 (2019).

Acknowledgements

Y.L. and E.A.K. acknowledge support by a New Frontier Grant from Cornell University's College of Arts and Sciences. E.A.K. acknowledges support by the National Science Foundation under grant no. OAC-2118310, the Ewha Frontier 10-10 Research Grant and the Simons Fellowship in Theoretical Physics award no. 920665. E.A.K. performed a part of this work at the Aspen Center for Physics, which is supported by the National Science Foundation grant no. PHY-160761.

Author information

Authors and Affiliations

1. Google Research, Mountain View, CA, USA

   T. I. Andersen, K. Kechedzhi, I. K. Drozdov, A. Bengtsson, S. Hong, A. Morvan, X. Mi, A. Opremcak, R. Acharya, R. Allen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B.

B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Z. Chen, B. Chiaro, D. Chik, C. Chou, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. Del Toro Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, V. S. Ferreira, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, A. G. Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, J. Hilton, M. R. Hoffmann, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. W. Lee, B. J. Lester, A. T. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, O. Martin, J. R. McClean, T. McCourt, M. McEwen, K. C. Miao, A. Mieszala, M. Mohseni, S. Montazeri, E. Mount, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, S. Omonije, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skruzny, W. C. Smith, R. Somma, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, H. Neven, S. Boixo, A. Megrant, J. Kelly, Y. Chen, V. Smelyanskiy, I. Aleiner & P. Roushan

2. Department of Physics, Cornell University, Ithaca, NY, USA

   Y. D. Lensky & E.-A. Kim

3. Department of Physics, University of Connecticut, Storrs, CT, USA

   I. K. Drozdov

4. Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA

J. C. Bardin

5. Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA

M. C. Hamilton

6. QSI, Faculty of Engineering & Information Technology, University of Technology Sydney, Sydney, New South Wales, Australia

M. Kieferová

7. Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA

A. N. Korotkov

8. Department of Chemistry, Columbia University, New York, NY, USA

J. Lee

9. Department of Physics, University of California, Santa Barbara, CA, USA

M. McEwen

10. Department of Physics and Astronomy, University of California, Riverside, CA, USA

L. P. Pryadko

11. Department of Physics, Ewha Womans University, Seoul, South Korea

E.-A. Kim

12. Department of Physics, Harvard University, Cambridge, MA, USA

E.-A. Kim

13. Radcliffe Institute for Advanced Studies, Cambridge, MA, USA

E.-A. Kim

Consortia

Google Quantum AI and Collaborators

- T. I. Andersen
- , Y. D. Lensky
- , K. Kechedzhi
- , I. K. Drozdov
- , A. Bengtsson
- , S. Hong
- , A. Morvan
- , X. Mi
- , A. Opremcak
- , R. Acharya
- , R. Allen
- , M. Ansmann
- , F. Arute
- , K. Arya
- , A. Asfaw
- , J. Atalaya
- , R. Babbush
- , D. Bacon
- , J. C. Bardin
- , G. Bortoli
- , A. Bourassa
- , J. Bovaird
- , L. Brill
- , M. Broughton
- , B. B. Buckley
- , D. A. Buell
- , T. Burger
- , B. Burkett
- , N. Bushnell
- , Z. Chen
- , B. Chiaro

- , D. Chik
- , C. Chou
- , J. Cogan
- , R. Collins
- , P. Conner
- , W. Courtney
- , A. L. Crook
- , B. Curtin
- , D. M. Debroy
- , A. Del Toro Barba
- , S. Demura
- , A. Dunsworth
- , D. Eppens
- , C. Erickson
- , L. Faoro
- , E. Farhi
- , R. Fatemi
- , V. S. Ferreira
- , L. F. Burgos
- , E. Forati
- , A. G. Fowler
- , B. Foxen
- , W. Giang
- , C. Gidney
- , D. Gilboa
- , M. Giustina
- , R. Gosula
- , A. G. Dau
- , J. A. Gross
- , S. Habegger
- , M. C. Hamilton
- , M. Hansen
- , M. P. Harrigan
- , S. D. Harrington
- , P. Heu
- , J. Hilton
- , M. R. Hoffmann

- , T. Huang
- , A. Huff
- , W. J. Huggins
- , L. B. Ioffe
- , S. V. Isakov
- , J. Iveland
- , E. Jeffrey
- , Z. Jiang
- , C. Jones
- , P. Juhas
- , D. Kafri
- , T. Khattar
- , M. Khezri
- , M. Kieferová
- , S. Kim
- , A. Kitaev
- , P. V. Klimov
- , A. R. Klots
- , A. N. Korotkov
- , F. Kostritsa
- , J. M. Kreikebaum
- , D. Landhuis
- , P. Laptev
- , K.-M. Lau
- , L. Laws
- , J. Lee
- , K. W. Lee
- , B. J. Lester
- , A. T. Lill
- , W. Liu
- , A. Locharla
- , E. Lucero
- , F. D. Malone
- , O. Martin
- , J. R. McClean
- , T. McCourt
- , M. McEwen

- , K. C. Miao
- , A. Mieszala
- , M. Mohseni
- , S. Montazeri
- , E. Mount
- , R. Movassagh
- , W. Mruczkiewicz
- , O. Naaman
- , M. Neeley
- , C. Neill
- , A. Nersisyan
- , M. Newman
- , J. H. Ng
- , A. Nguyen
- , M. Nguyen
- , M. Y. Niu
- , T. E. O’Brien
- , S. Omonije
- , A. Petukhov
- , R. Potter
- , L. P. Pryadko
- , C. Quintana
- , C. Rocque
- , N. C. Rubin
- , N. Saei
- , D. Sank
- , K. Sankaragomathi
- , K. J. Satzinger
- , H. F. Schurkus
- , C. Schuster
- , M. J. Shearn
- , A. Shorter
- , N. Shutty
- , V. Shvarts
- , J. Skruzny
- , W. C. Smith
- , R. Somma

- , G. Sterling
- , D. Strain
- , M. Szalay
- , A. Torres
- , G. Vidal
- , B. Villalonga
- , C. V. Heidweiller
- , T. White
- , B. W. K. Woo
- , C. Xing
- , Z. J. Yao
- , P. Yeh
- , J. Yoo
- , G. Young
- , A. Zalcman
- , Y. Zhang
- , N. Zhu
- , N. Zobrist
- , H. Neven
- , S. Boixo
- , A. Megrant
- , J. Kelly
- , Y. Chen
- , V. Smelyanskiy
- , E.-A. Kim
- , I. Aleiner
- & P. Roushan

Contributions

Y.D.L., K.K., E.-A.K. and I.A. developed the underlying theory. T.I.A., Y.D.L., K.K., E.-A.K., I.A. and P.R. developed the experiment. T.I.A. performed the experiment and analysed the data. I.K.D., A.B., S.H., A.M., X.M. and A.O. provided assistance with calibration. T.I.A., Y.D.L., K.K., E.-A.K., I.A. and P.R. wrote the manuscript. T.I.A., Y.D.L., K.K. and P.R. wrote the Methods. All authors contributed to revising the manuscript. All authors

contributed to the experimental and theoretical infrastructure to enable the experiment.

Corresponding authors

Correspondence to E.-A. Kim, I. Aleiner or P. Roushan.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Bernard van Heck and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Qubit relaxation (T_1) and coherence (T_2) times.

a,**b**, Cummulative distributions of T_1 (**a**) and T_2 (**b**), where the latter is measured using a Hahn echo sequence. Dashed lines indicate the median

values of 21.7 μs for both measures. Insets: T_1 and T_2 plotted against qubit number.

Extended Data Fig. 2 Gate errors.

a,**b**, Cummulative distributions of the Pauli error for single-qubit (**a**) and two-qubit CZ (**b**) gates. We find median error values of 1.3×10^{-3} and 7.3×10^{-3} for the single-qubit and CZ gates, respectively.

Extended Data Fig. 3 Readout benchmarking and correction.

a, Histogram of readout error, with a median value of 2.0% (dashed vertical line). Inset: readout error plotted against qubit number. **b**,**c**, Stabilizer values of the surface code ground state before (**b**) and after (**c**) readout correction.

Extended Data Fig. 4 Dynamical decoupling.

a,**b**, Stabilizer values without (**a**) and with (**b**) dynamical decoupling, after D3V braiding. Dynamical decoupling improves the average absolute stabilizer value from 0.50 to 0.58.

Extended Data Fig. 5 Circuit details.

Circuits used for the fusion experiment (**a**), the full-braid experiment (**b**), and the half-braid experiment (**c**), shown in Figs. 2–4, respectively, in the main text. Turqoise and gray boxes denote dynamical decoupling and phased *XZ*-gates, respectively. In the full-braid experiment (**b**), we include five single-qubit rotations to permute \(\hat{X}\), \(\hat{Y}\) and \(\hat{Z}\) of the three stabilizers touching the moving D3V in steps V-VIII and IX-XI, as well as three Hadamard-gates to return all stabilizers to the original \(\hat{Z}\hat{X}\hat{X}\hat{Z}\)-form in XII. See Extended Data Fig. 6 for the circuit used for ground state preparation, as well as details on how the multi-qubit unitary gates used to move anyons are decomposed into CZ-gates.

Extended Data Fig. 6 Ground state preparation and CZ-decompositions.

a, Schematic showing the circuit used for preparation of the ground state of the surface code. The protocol is the same as that shown in ref. [43], except with the inclusion of Hadamard gates on alternating qubits in the final step, since we use symmetrized stabilizers on the form \(\hat{Z}\hat{X}\hat{X}\hat{Z}\). **b**, The unitary needed to move a D3V between two neighboring vertices is realized in the experiment through the use of one CZ-gate and single-qubit rotations. **c**, When D3Vs are moved diagonally, we include two SWAP gates, requiring three CZ-gates each. **d**, Main: the three-qubit unitary used in step VIII in Fig. 3 is equivalent to a combination of single-qubit gates, 4 SWAP-gates and 4 CZ-gates. Right dashed box: decomposition of a SWAP-gate into CZ-gates. **e**, Adjacent single-qubit gates are merged and shifted left to the nearest CZ-gate.

Extended Data Fig. 7 Simulation of braiding in the presence of noise.

a, Simulation results. **b**, Experimental data (same as in step XII in Fig. 3b). We observe relatively good agreement between the simulation and the experimental results, except some discrepancies that are attributed to inhomogeneity of the errors.

Extended Data Fig. 8 Braiding distinguishable D3Vs.

a, Braiding schematic of worldlines. **b**, Step-by-step depiction of stabilizers as the two σ are braided, analogous to that in Fig. 3, but with distinguishable σ.

Extended Data Fig. 9 Alternative protocol for braiding σ.

a, Schematic displaying the braiding process of the two σ-pairs. **b**, Experimental demonstration of braiding, displaying the values of the stabilizers throughout the process. Two σ-pairs, A and B, are created from the vacuum \({\mathbb{1}}\), and one of the D3Vs in pair A is brought to

the right side of the grid. Next, a σ from pair B is moved to the top, thus crossing the path of the first σ, before bringing the σ from pair A back again to complete the braid. The diagonal σ move performed in step VI is achieved by including two SWAP-gates, corresponding to 6 additional CZ-gates. The yellow triangles represent the locations of the σ, while the brown and green lines represent the paths of σ from pair A and B, respectively. The average absolute stabilizer value is 0.93 ± 0.06 and 0.77 ± 0.09 in the first and last step, respectively. **c**, After braiding the σ, we search for hidden fermions by measuring the Pauli string $\hat{{\mathcal{P}}}$ (left panels), which here is equivalent to $\hat{Y}$ on the qubit where the two σ overlap. The measurement yields $\langle \hat{{\mathcal{P}}}\rangle =\langle \hat{Y}\rangle =-\,0.71\,\pm \,0.01$, indicating creation of a fermion. Right: world-lines of braiding process, including non-local measurement based on plaquette violation loop. **d**, Same as **c**, but after braiding two distinguishable σ, achieved by applying the inverse two-qubit gates when moving the σ in pair B. The measurement yields $\langle \hat{Y}\rangle =+\,0.71\,\pm \,0.01$, indicating no fermion creation.

Supplementary information

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Google Quantum AI and Collaborators. Non-Abelian braiding of graph vertices in a superconducting processor. *Nature* **618**, 264–269 (2023). https://doi.org/10.1038/s41586-023-05954-4

- Received: 03 October 2022
- Accepted: 14 March 2023
- Published: 11 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05954-4

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

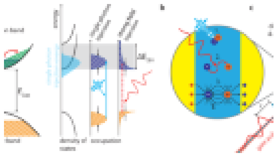
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

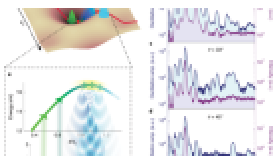
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05954-4



The speed limit of optoelectronics

25 March 2022

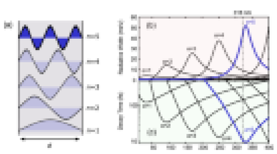
M. Ossiander, K. Golyari, … M. Schultze



Attosecond spectral singularities in solid-state high-harmonic generation

20 January 2020

Ayelet Julie Uzan, Gal Orenstein, … Nirit Dudovich



Ultrafast thermal-free photoluminescence of coherently extended single quantum states

11 June 2019

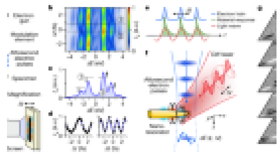
Takuya Matsuda, Masayoshi Ichimiya, … Hajime Ishihara



[Absolute timing of the photoelectric effect]

19 September 2018

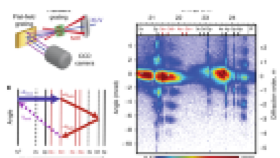
M. Ossiander, J. Riemensberger, … R. Kienberger



[Attosecond electron microscopy of sub-cycle optical dynamics]

31 May 2023

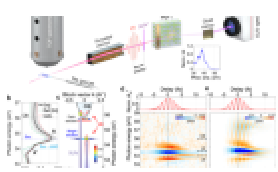
David Nabben, Joel Kuttruff, … Peter Baum



[Nonlinear XUV signal generation probed by transient grating spectroscopy with attosecond pulses]

27 March 2019

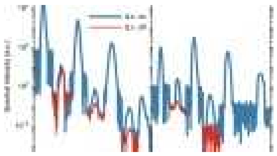
Ashley P. Fidler, Seth J. Camp, … Stephen R. Leone



[Unravelling the intertwined atomic and bulk nature of localised excitons by attosecond spectroscopy]

15 February 2021

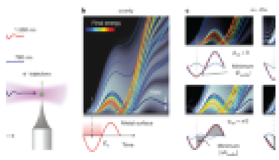
Matteo Lucchini, Shunsuke A. Sato, … Mauro Nisoli



[**Strong-field optoelectronics in solids**]

18 June 2018

G. Vampa, T. J. Hammond, … P. B. Corkum



[**Tracing attosecond electron emission from a nanometric metal tip**]

26 April 2023

Philip Dienstbier, Lennart Seiffert, … Peter Hommelhoff

- Article
- Open Access
- Published: 24 May 2023

Dynamic optical response of solids following 1-fs-scale photoinjection

- Dmitry A. Zimin,
- Nicholas Karpowicz ,
- Muhammad Qasim,
- Matthew Weidman,
- Ferenc Krausz &
- …
- Vladislav S. Yakovlev

Show authors

Nature **volume 618**, pages 276–280 (2023)

- 2829 Accesses
- 77 Altmetric
- Metrics details

Abstract

Photoinjection of charge carriers profoundly changes the properties of a solid. This manipulation enables ultrafast measurements, such as electric-field sampling[1,2], advanced recently to petahertz frequencies[3,4,5,6,7], and the real-time study of many-body physics[8,9,10,11,12,13]. Nonlinear photoexcitation by a few-cycle laser pulse can be confined to its strongest half-cycle[14,15,16]. Describing the associated subcycle optical response, vital for attosecond-scale optoelectronics, is elusive when studied with traditional pump-probe metrology as the dynamics distort any probing field on the timescale of the carrier, rather than that of the envelope. Here we apply field-resolved optical metrology to these dynamics and report the direct observation of the evolving optical properties of silicon and silica during the first few femtoseconds following a near-1-fs carrier injection. We observe that the Drude–Lorentz response forms within several femtoseconds—a time interval much shorter than the inverse plasma frequency. This is in contrast to previous measurements in the terahertz domain[8,9] and central to the quest to speed up electron-based signal processing.

Main

The nonlinear interaction of an intense laser pulse with a medium changes its refractive index. These changes may disappear immediately after the end of the pulse, as in the case of the optical Kerr effect. They may also outlast the pulse if electrons are promoted to excited states. Notably, nonlinear processes are confined to a time interval shorter than the pulse duration,

multiphoton excitation being a prominent example[17]. Such a nonlinearity allows the shortest light pulses, with a single dominant half-cycle of the electric field, to photoinject most charge carriers within a fraction of that half-cycle. The resultant extreme temporal confinement of carrier injection opens new opportunities for ultrafast science[18,19,20,21,22,23].

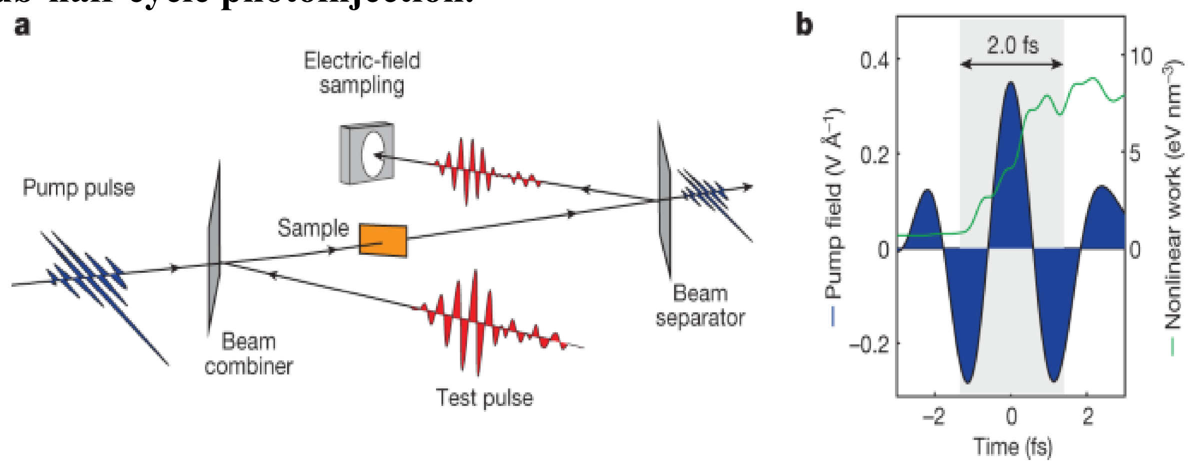
Here we investigate how the optical response of this electron–hole plasma forms after sudden, 1-fs-scale, photoexcitation of valence electrons. The time required for the response to build up depends not only on the duration of the laser pulse that creates the plasma, but also on how long it takes quasiparticles to acquire their properties. Collective behaviour, such as Coulomb screening and plasma scattering, is expected to form on the timescale of the inverse plasma frequency, henceforth called the plasma period[8,24,25,26,27]. Thus, a question of far-reaching implications is whether the plasma frequency sets a fundamental speed limit for future advancement of optoelectronic signal processing and metrology. The time it takes the plasma response to emerge is also essential for modelling the interaction of solids with intense laser pulses of a few cycles. This highly nonlinear interaction may populate states in hundreds of energy bands, which presents formidable challenges for an ab initio description of many-electron dynamics. Thus, efficient and accurate modelling of the physical processes underlying cutting-edge metrology and signal processing critically depends on the role that many-body physics plays during the first few femtoseconds after photoexcitation.

Field-resolved pump-probe measurements

In our experiments, a 3 fs (full-width at intensity half-maximum) linearly polarized near-infrared pump pulse (with a period of carrier frequency of 2.3 fs) photoexcited silicon (with a direct band gap of 3.2 eV) and fused-silica (with an energy gap of 9 eV) by multiphoton absorption. The arrival time, τ, of this pump pulse was varied with respect to a weak 12 fs test pulse carried at a central wavelength of 2.1 μm (with a period of carrier frequency of 7 fs). For Si, photoinjection was confined to approximately 2 fs, as shown by the green curve in Fig. 1b, which represents the nonlinear work performed by the pump field. This time-dependent work is mostly spent on creating charge carriers, and we calculated its value using time-dependent

density functional theory[28]. The relatively long period of the test field is beneficial for studying light-driven electron motion. To decouple this motion from the photoinjection, the pump and test pulses were polarized orthogonally. The transmitted pump pulse was blocked using a wire-grid polarizer, while the oscillating electric field (henceforth, waveform) of the test pulse was recorded using the recently developed generalized heterodyne optical sampling technique (GHOST)[29].

Fig. 1: Field-resolved detection of charge-carrier motion controlled by sub-half-cycle photoinjection.



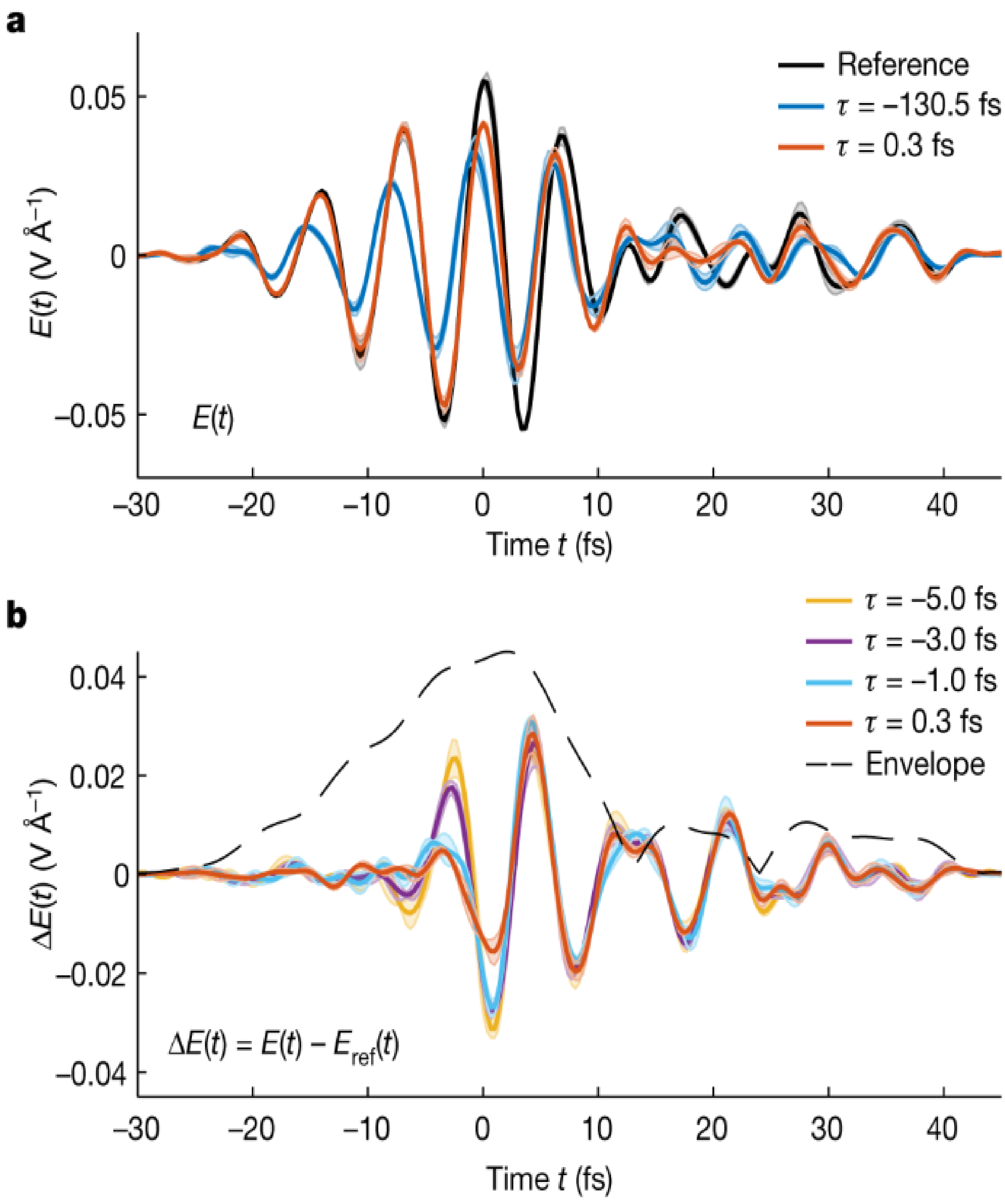
a, A schematic of the pump-probe measurements in which the waveform of the test pulse (red) was measured after transmission through a thin sample (orange) for different delays of the pump pulse (blue), which created charge carriers by multiphoton absorption. **b**, The nonlinear nature of the photoinjection confines its duration to a time interval shorter than the half wave-cycle of the test pulse, even for the weaker (two-photon) nonlinearity occurring in Si (shown in the panel). The green curve shows the nonlinear work performed by the pump field. The grey area highlights the time between the moments at which the work reaches 10% and 90% of its final value. In SiO_2, this confinement is even stronger owing to the higher number of photons needed for promoting a valence electron to the conduction band[16].

Source data

Figure 2 displays waveforms transmitted through a 230-nm-thick silicon sample. In Fig. 2a, we compare the reference waveform, $E_{ref}(t)$, which is the

transmitted test pulse without photoinjection, with the waveforms $E(t)$ recorded for two arrival times of the 0.8 V $Å^{-1}$ pump pulse: before and at the centre of the test pulse. In the latter case, the waveform remains unchanged until photoinjection. Figure 2b plots the change in the transmitted test field (with and without carrier injection) for different moments of injection. This change, $\Delta E(t) = E(t) - E_{ref}(t)$, builds up in the first half-cycle of the test field (that is, within 3–4 fs) following the peak of the pump pulse and starts decaying another half-cycle later. The experiment provides a textbook example for how the material polarization (in this case induced by the injected charge carriers) is imprinted in the electric field transmitted through the sample.

Fig. 2: The effect of ultrafast photoinjection on the test pulse transmitted through the 230-nm-thick silicon sample.



a, The reference waveform \({E}_{{\rm{ref}}}(t)\) (black), which is the electric field of the test pulse transmitted through the unperturbed sample, is compared with a test waveform with photoexcitation preceding the test pulse (blue) and the waveform measured with the 3 fs pump pulse arriving at $t = 0$ (orange). The shaded areas represent the standard deviations from three independent measurements. **b**, The measured photoexcitation-induced

change in the transmitted electric field, ΔE, for several arrival times of the pump pulse, τ. The dashed curve shows the envelope of \({E}_{{\rm{ref}}}(t)\).

Source data

The measured photoexcitation-induced waveform distortions carry information about how the optical thickness and absorption of the sample change with time. In the static case, these properties are fully described by a complex refractive index. In the Methods, we define a generalized refractive index that describes the medium's dynamic response. For a dispersive medium, such a generalized refractive index depends on both frequency, ω, and time, t. The formation of an electron–hole plasma decreases the real part of the refractive index, which increases the phase velocity of the test pulse and thus shifts its carrier wave to earlier times. The photoinjected electrons and holes also absorb light, which attenuates the test waveform transmitted after the pump pulse. We clearly see both these effects in Fig. 2a. From Fig. 2b we see that already 3–4 fs after the peak of the pump pulse the waveform distortions measured for slightly different delays merge within their standard deviations. We interpret this as a transition from ballistic to non-ballistic conduction.

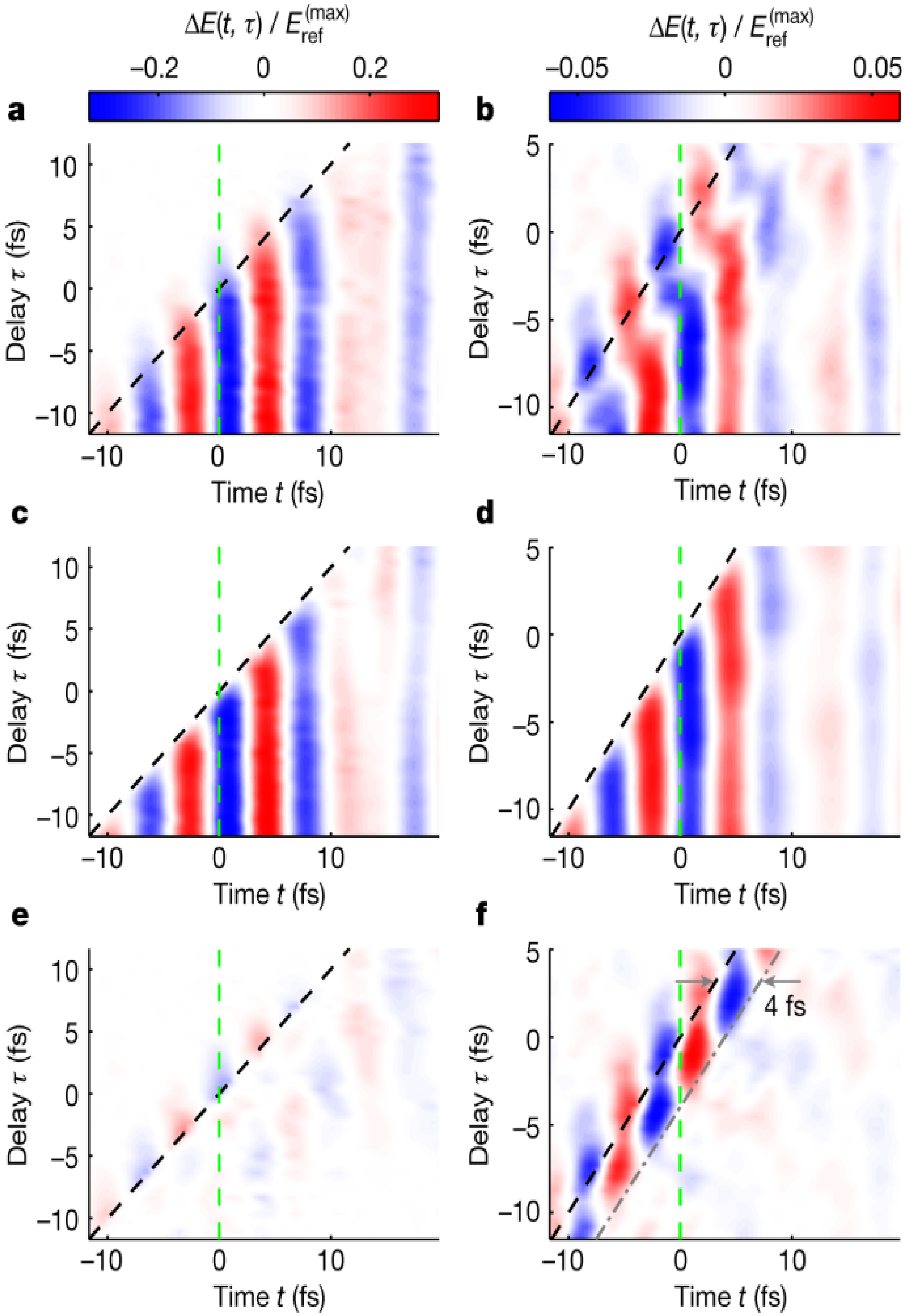
Drude–Lorentz model

For gaining insight into how ultrafast photoinjection reshapes the test pulse, it is sensible to seek the simplest possible model able to account for the essential physics. For this purpose, we use the Drude–Lorentz model in the idealized case of instantaneous photoinjection. A textbook derivation of the Drude–Lorentz model[30] considers the classical motion of free charge carriers and bound electrons in an oscillating electric field. In the Methods, we summarize and state again this derivation for the specific case in which pairs of charge carriers are created and begin their classical motion under the influence of the electric force of the test field at a certain moment τ (similar models were previously developed for terahertz time-domain spectroscopy[31,32,33]). The result is a complex time-dependent refractive index that describes the properties of the electron–hole plasma but does not account for wave-mixing processes that require the presence of the pump

pulse, such as the optical Kerr effect or the dynamical Franz–Keldysh effect[34,35]. On the basis of this time-dependent refractive index, we investigate the extent to which the observed waveform distortion can be accounted for by the Drude–Lorentz model.

To this end, we numerically optimize the phenomenological parameters of the model to fit the measured waveform distortions in the time interval $t \geq \tau + 5$ fs. In this time window, we can neglect any nonlinear interaction of the 3 fs pump pulse with charge carriers that could induce a polarization response in the direction of the test field. The results are shown in Fig. 3, which compares the measurements with the 230-nm-thick silicon sample to those with a 12.7-μm-thick fused-silica sample, for which the pump pulse had a peak field strength of 1.4 V Å^{-1}. In Fig. 3e,f, we plot the difference between the measured and reconstructed waveform distortions. For both materials, the model accurately reconstructs the measured delay-dependent waveform distortion except in a time interval of few femtoseconds around the centre of the pump pulse. In the fused-silica measurements (Fig. 3f), we evaluate an upper limit for the formation time of the Drude–Lorentz response to be 4 fs. This is notable given that the plasma period, which we retrieve from these measurements, is 110 fs. For silicon, we estimate the plasma period to be between 2 and 3 fs, which is comparable to the duration of carrier injection.

Fig. 3: Distortion of the test waveforms induced by the pump pulse.



ΔE(t, τ) / E_ref^(max)
−0.2
0
0.2
a
Delay τ (fs)
Time t (fs)
ΔE(t, τ) / E_ref^(max)
−0.05
0
0.05
b
Delay τ (fs)
Time t (fs)
c
Delay τ (fs)
Time t (fs)
d
Delay τ (fs)
Time t (fs)
e
Delay τ (fs)
Time t (fs)
f
4 fs
Delay τ (fs)
Time t (fs)

These pseudocolor diagrams show $\Delta E(t,\tau \,)$ normalized to the peak value of the reference field, ${E}_{{\rm{ref}}}^{({\rm{\max }})}$. The black dashed lines show the middle of the pump pulse ($t=\tau$). The green dashed lines represent $t = 0$, at which the reference waveform reaches its maximal value. **a**,**b**, $\Delta E(t,\tau \,)$ measured for silicon (**a**) and silica (**b**). **c**,**d**, $\Delta E(t,\tau \,)$ reconstructed using the Drude–Lorentz model for silicon (**c**) and silica (**d**). **e**,**f**, By subtracting the reconstructed $\Delta E\left(t,\tau \,)\right)$ from the measured one, we obtain the non-Drude–Lorentz waveform distortions, which are shown for silicon (**e**) and silica (**f**), using the same colour schemes as those in the upper panels.

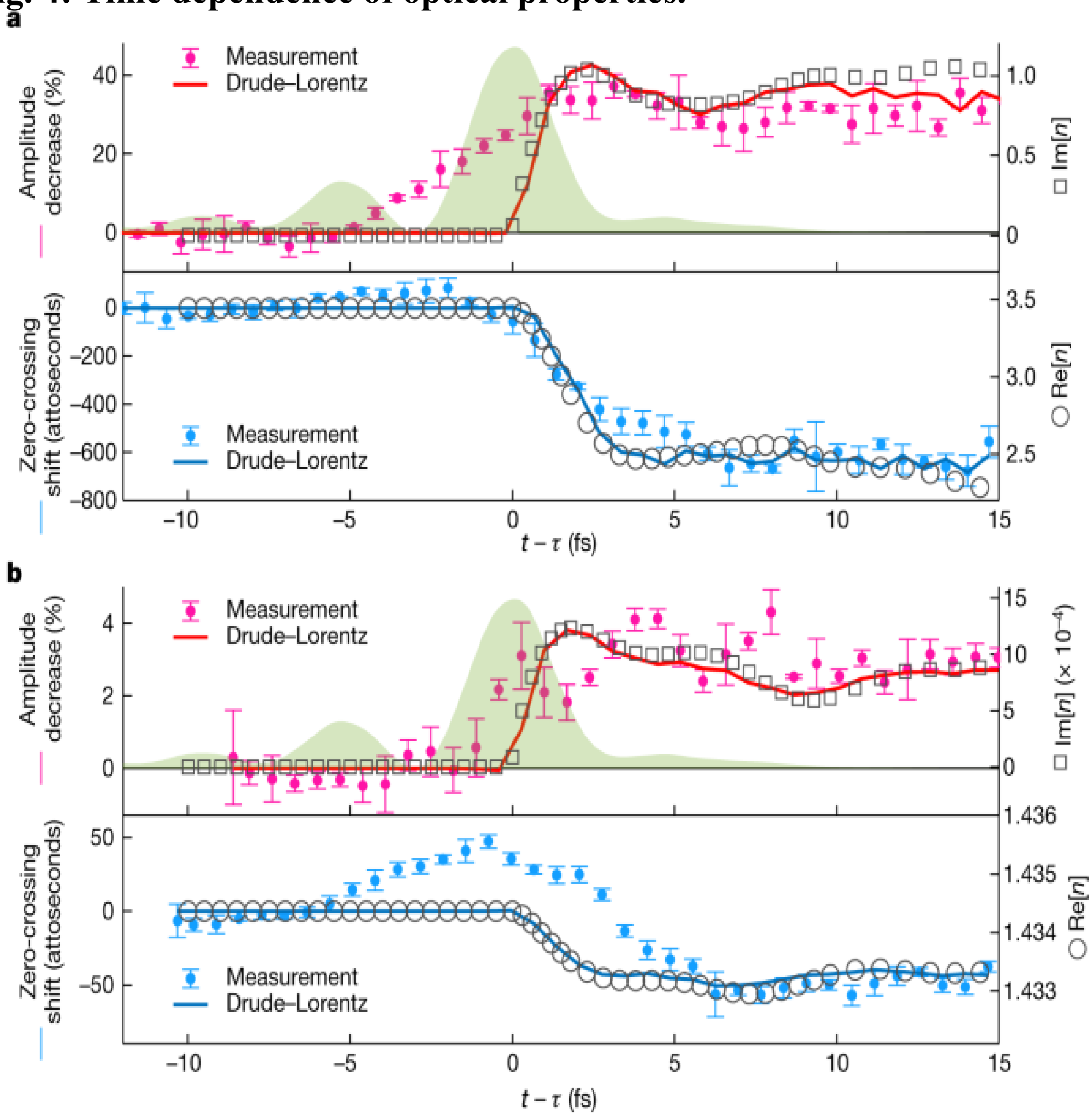
Time-dependent refractive index

Whereas the interaction with the pump pulse is nonlinear, the polarization response to the weak test pulse is linear. So, its transmission through the sample is fully described by the complex-valued, time-dependent refractive index ${n}_{\omega }(t)$, which does not depend on the shape of the test waveform and encodes all the information about the physical processes that make the medium's optical properties evolve. It is possible to retrieve the time-dependent refractive index from the waveform distortions. We find, however, that a good first approximation to the time-dependent refractive index at the central frequency of the test pulse may be obtained without sophisticated reconstruction by applying the following analysis to delay-dependent waveforms.

A mere change in the optical thickness of a sample barely affects half-cycle amplitudes but it shifts the zero crossings of the electric field in time. At the same time, changes in absorption and reflection strongly affect the half-cycle amplitudes, but not the positions of the field's zero crossings. This indicates that the delay dependence of zero-crossing shifts should be closely related to Re[${n}_{\omega }(t)$], whereas the delay dependence of the half-cycle amplitudes should be mainly determined by Im[${n}_{\omega }(t)$]. According to our modelling, this indeed is the case in our measurements. The error bars in Fig. 4 show how photoinjection shifts the zero crossings and reduces the half-cycle amplitudes in the measured data. These error bars represent standard deviations evaluated from delay scans that were

performed one after another (three scans for silicon and five scans for fused silica).

Fig. 4: Time dependence of optical properties.



a, The analysis of the silicon data. The positions of the magenta error bars and the red curve show the photoinjection-induced decrease of the half-cycle amplitudes in the measured and reconstructed waveforms, respectively. These data are juxtaposed with the imaginary part of the refractive index (black squares), evaluated using the Drude–Lorentz model. The positions of the light-blue error bars and the blue curve represent the change in the timing of the test field’s zero crossings in the measured and reconstructed

waveforms, respectively. The black circles show the real part of the time-dependent refractive index. **b**, The same analysis applied to the fused-silica data. The horizontal axis in these plots is the time passed after the middle of the pump pulse. The intensity envelope of the pulse is shown by the green area.

Source data

The solid curves in Fig. 4 were obtained by applying the same analysis to the reconstructed waveforms, for which we used the Drude–Lorentz model with the same parameters as before (in Fig. 3). The real and imaginary parts of the refractive index in this model are depicted with empty circles and squares, respectively. More accurately, this time-dependent refractive index describes the propagation of a monochromatic test wave with a wavelength of 2.1 μm, which is the central wavelength of the test pulse. We see that the shape of Re[$n_{\omega}(t)$] closely resembles that of the zero-crossing shift, whereas the shape of Im[$n_{\omega}(t)$] matches that of the half-cycle amplitude reduction. We also observe here that neither the theoretical refractive index nor the reconstructed waveforms change abruptly even though photoinjection was modelled as an instantaneous event. This is because, in such measurements, $\Delta E(t)$ is approximately proportional to the electric current induced by the test field[36], and the current gradually builds up as the field accelerates the charge carriers. For the same reason, the time dependence of $n_{\omega}(t)$ does not end with photoinjection (see equation (16) in the Methods).

With these insights, Fig. 4 allows us to further analyse the part of the medium response that is not accounted for by our Drude–Lorentz model with instantaneous photoinjection. In the case of silicon (Fig. 4a), the modelled change in the refractive index is steeper than that inferred from the measurements: measured between 10% and 90% of the maximum change, the rise time of absorption is 5.2 fs in the measurements and 1.4 fs in the simulations. This discrepancy can be largely explained by the approximation of instantaneous photoinjection that we made in our model. Even if we neglect phonon-assisted single-photon transitions, we expect that 80% of the charge carriers are photoinjected within a time interval of approximately 2 fs (Fig. 1b). We conclude that the results in Fig. 4a do not contain any

unambiguous evidence for many-body dynamics prolonging the buildup of the Drude–Lorentz response.

In the case of fused silica (Fig. 4b), photoinjection requires the absorption of at least four photons from the pump pulse. Hence, we expect that more than 70% of the charge carriers appeared during the central half-cycle of the pulse. Here the approximation of instantaneous photoinjection is more applicable, and other effects are responsible for the differences between the error bars and the solid curves. The test field’s zero crossings first get delayed and only then experience a negative time shift. We associate this transient increase in the optical thickness of the sample with the optical Kerr effect. We cannot fully explain why the positive zero-crossing shifts extend over 8 fs, which is three times as large as the full-width half-maximum of the pump pulse, but the propagation through the 12.7 μm sample may be partially responsible for it. We had to use a relatively thick fused-silica sample to achieve an acceptable signal-to-noise ratio in spite of the miniscule changes in the refractive index—the relative inefficiency of the high-order multiphoton absorption in fused silica puts an upper limit on the concentration of charge carriers that can be achieved without destroying the sample. In contrast to silicon, the contribution from the electron–hole plasma to Δn in fused silica was comparable in magnitude to that from wave-mixing processes enabled by the presence of the pump pulse. In addition to the optical Kerr effect, these processes include a transient increase in absorption, which is a manifestation of the Franz–Keldysh effect (the Keldysh parameter was close to 1 in the fused-silica measurements).

Conclusions

In summary, petahertz-scale optical-field metrology in a pump-probe setting enables the direct observation of how the optical properties of a medium evolve after 1-fs-scale photoinjection. For a sufficiently thin sample, the time dependence of the real part of the refractive index is closely matched by the delay dependence of the test waveform’s zero crossings, whereas the imaginary part of the refractive index matches the delay dependence of the half-cycle amplitudes.

Both in silicon and in fused silica, we observed that the Drude–Lorentz response forms within a few femtoseconds after photoinjection. In fused silica, this time was much shorter than the inverse plasma frequency. Under the premise that the plasma frequency sets the relevant timescale for the formation of charge–charge interactions in many-body systems, we may conclude that many-body phenomena have little effect on how a photoexcited wide-gap material, in this case fused silica, responds to visible or infrared light. This conclusion grants credibility to treating the interaction of abruptly injected carriers with optical fields in theoretical frameworks in which electrons are described with single-particle mathematical objects[37,38], for example by means of time-dependent density functional theory[39] or semiconductor Bloch equations[40]. At the same time, it is likely that future research will find examples in which collective processes are pronounced in suddenly photoexcited solids and petahertz-scale optical-field metrology will be a powerful tool for investigating them.

Methods

The linear polarization response of a medium with time-dependent properties

When a pulse of light is transmitted through a material, the textbook description of the interaction between the electric field of the pulse and matter is generally given in terms of material parameters, such as the refractive index, dielectric function or susceptibility. When the material changes its properties slowly, the generalization of these quantities is straightforward. A more careful generalization is necessary when the properties substantially change within a single oscillation of the optical field. This is the purpose of this and the following sections.

Let us consider the polarization response induced by a light pulse that is so weak that the relationship between its electric field, $E(t)$, and the induced polarization, $P(t)$, can be assumed to be linear. By definition, a linear response must satisfy two requirements. First, if $E(t)$ induces $P(t)$, then $aE(t)$ must induce $aP(t)$, where a is a constant. Second, if ${E}_{1}\left(t\right)$ induces ${P}_{1}\left(t\right)$ and ${E}_{2}\left(t\right)$ induces \

$P_{2}\left(t\right)$, then the sum of the fields, $E_{1}\left(t\right)+E_{2}\left(t\right)$, must induce the polarization that is equal to $P_{1}\left(t\right)+P_{2}\left(t\right)$. For simplicity, we regard the electric field and the induced polarization as scalar quantities and we assume that the polarization response is local. When the medium is static, the most general expression of the linear polarization response that satisfies causality is given by the well-known equation

$$P\left(t\right)={\int }_{0}^{{\rm{\infty }}}{\rm{d}}\tau \,{\chi }^{\left(1\right)}\left(\tau \right)E\left(t-\tau \right),$$

which we have written in CGS units.

Here ${\chi }^{\left(1\right)}\left(\tau \right)$ is the first-order, time-domain susceptibility. As the right-hand side of the above expression is a convolution, the linear polarization response of a static medium looks particularly simple in the frequency domain: $\widetilde{P}\left(\omega \right)={\chi }_{\omega }^{\left(1\right)}\widetilde{E}\left(\omega \right)$.

Let us now generalize this formalism to the case in which the optical properties of the medium evolve in time. The time-domain description is a convenient starting point for such a generalization.

The following equation satisfies the causality principle and the linearity requirements:

$$P\left(t\right)={\int }_{0}^{{\rm{\infty }}}{\rm{d}}\tau \,{\chi }^{\left(1\right)}\left(\tau ,t\right)E\left(t-\tau \right).$$
(1)

The linear susceptibility now depends on two arguments. The first one, τ, is required because the polarization response does not build up instantaneously. The second argument, t, represents the time dependence of the medium’s properties. Because of this second argument, the Fourier transform of the integral is no longer a product of $E(t)$ and a frequency-domain susceptibility; one must also perform a convolution when working in the frequency domain. This means that a medium with time-dependent properties can mix

frequency components of *E*(*t*) and generate new ones. At this point, it is instructive to consider two special cases: the impulse response and the response to a monochromatic wave. These cases are useful because any pulse can be written as a linear superposition of delta spikes,

$$E\left(t\right)={\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{{\rm{d}}t}_{0}\,E\left({t}_{0}\right)\delta \left(t-{t}_{0}\right)$$
(2)

or monochromatic waves,

$$E\left(t\right)=\frac{1}{2{\rm{\pi }}}{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}\omega \,\widetilde{E}\left(\omega \right){{\rm{e}}}^{-{\rm{i}}\omega t},$$
(3)

the latter equation being just the inverse Fourier transform. By substituting $E\left(t\right)$ with $\delta \left(t-{t}_{0}\right)$ in equation (1), we get the polarization response to a delta spike that arrives at time ${t}_{0}$:

$${P}_{\delta }\left(t,{t}_{0}\right)={\int }_{0}^{{\rm{\infty }}}{\rm{d}}\tau \,{\chi }^{\left(1\right)}\left(\tau ,t\right)\delta \left(t-\tau -{t}_{0}\right)={\chi }^{\left(1\right)}\left(t-{t}_{0},t\right)\varTheta \left(t-{t}_{0}\right).$$
(4)

Here $\varTheta $ is the Heaviside function, which can be incorporated into the definition of ${\chi }^{\left(1\right)}\left(\tau ,t\right)$. We can rewrite this equation in the form that gives the recipe for evaluating ${\chi }^{\left(1\right)}\left(\tau ,t\right)$ from the polarization response induced by a delta spike:

$${\chi }^{\left(1\right)}\left(\tau ,t\right)={P}_{\delta }\left(t,t-\tau \right).$$
(5)

We see that ${\chi }^{\left(1\right)}\left(\tau ,t\right)$ is simply the polarization that the delta spike arriving at time $t-\tau $ induces at time

(t\).

Let us now return to equation (1) and substitute $E\left(t\right)$ with ${{\rm{e}}}^{-{\rm{i}}\omega t}$. This yields the response to a monochromatic wave:

$${P}_{\omega }\left(t\right)={\int }_{0}^{{\rm{\infty }}}{\rm{d}}\tau \,{\chi }^{\left(1\right)}\left(\tau ,t\right){{\rm{e}}}^{-{\rm{i}}\omega \left(t-\tau \right)}={{\rm{e}}}^{-{\rm{i}}\omega t}{\chi }_{\omega }\left(t\right),$$
(6)

where ${\chi }_{\omega }\left(t\right)$ is the Fourier transform of ${\chi }^{\left(1\right)}\left(\tau ,t\right)$ with respect to the first argument:

$${\chi }_{\omega }\left(t\right)={\int }_{0}^{{\rm{\infty }}}{\rm{d}}\tau \,{\chi }^{\left(1\right)}\left(\tau ,t\right){{\rm{e}}}^{{\rm{i}}\omega \tau }.$$
(7)

We can regard ${\chi }_{\omega }\left(t\right)$ as a time-dependent linear susceptibility to a monochromatic wave.

Equipped with ${\chi }_{\omega }\left(t\right)$, we can evaluate the linear polarization response to an arbitrary pulse as

$$P\left(t\right)=\frac{1}{2{\rm{\pi }}}{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}\omega \,\widetilde{E}\left(\omega \right){{\rm{e}}}^{-{\rm{i}}\omega t}{\chi }_{\omega }\left(t\right)$$
(8)

In the frequency domain, this translates into

$$\widetilde{P}\left(\omega \right)=\frac{1}{2\pi }{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}d{\omega }^{{\prime} }\widetilde{E}\left({\omega }^{{\prime} }\right){\widetilde{\chi }}_{{\omega }^{{\prime} }}\left(\omega -{\omega }^{{\prime} }\right),$$

where

$${\widetilde{\chi }}_{{\omega }^{{\prime} }}\left(\omega \right)={\int }_{\infty }^{-{\rm{\infty }}}{\rm{d}}t{\chi }_{{\omega }^{{\prime} }}\left(t\right){{\rm{e}}}^{{\rm{i}}\omega t}={\iint }_{-\infty }^{\infty }{\rm{d}}t{\rm{d}}\tau {\chi }^{\left(1\right)}\left(\tau ,t\right){{\rm{e}}}^{{\rm{i}}\left(\omega t+{\omega }^{{\prime} }\tau \right)}$$

is the two-dimensional Fourier transform of \({\chi }^{\left(1\right)}\left(\tau ,t\right)\) with respect to both arguments.

Time-dependent refractive index

For a static medium, the complex-valued refractive index, \({n}_{\omega }\), describes the propagation of a monochromatic wave according to the following formula:

$${E}_{\omega }\left(z,t\right)=\text{Re}\left[\exp \left({\rm{i}}\frac{\omega }{c}{n}_{\omega }z\right){{\rm{e}}}^{-{\rm{i}}\omega t}{E}_{\omega }\left(0,t\right)\right],$$

which is consistent with Maxwell's equations as long as c is the vacuum speed of light and \({n}_{\omega }^{2}=1+4{\rm{\pi }}{\chi }_{\omega }^{\left(1\right)}\) (for a non-magnetic medium). Although it may look reasonable to define a time-dependent refractive index as \(\sqrt{1+4\pi {\chi }_{\omega }^{\left(1\right)}\left(t\right)}\), this will not necessarily preserve the above expression for pulse propagation. Let us search for a definition of \({n}_{\omega }\left(t\right)\) that allows us to associate its real and imaginary parts with the optical thickness and the absorption of a thin film. The complex-valued transmittivity, \({T}_{\omega }\), of a thin sample relates the incident and transmitted fields: if the incident field is a monochromatic wave \({E}_{\omega }^{\left(0\right)}{{\rm{e}}}^{-{\rm{i}}\omega t}\), then the transmitted field is \({T}_{\omega }{E}_{\omega }^{\left(0\right)}{{\rm{e}}}^{-{\rm{i}}\omega t}\). Here the generalization to media with time-dependent optical properties is unambiguous:

$${E}_{\omega }^{\text{out}}\left(t\right)={T}_{\omega }\left(t\right){E}_{\omega }^{\text{in}}\left(t\right)={T}_{\omega }\left(t\right){E}_{\omega }^{\left(0\right)}{{\rm{e}}}^{-{\rm{i}}\omega t}.$$

(9)

If ${T}_{\omega }$ depends on time, ${E}_{\omega }^{\text{out}}\left(t\right)$ will no longer be a monochromatic wave—the *ω* subscript merely indicates the frequency of the incident wave. As long as the transmission is linear, knowing ${T}_{\omega }\left(t\right)$ allows one to evaluate the transmission of an arbitrary light pulse by decomposing it into monochromatic waves, transmitting each of them separately and assembling the outcomes:

$${E}^{\text{out}}\left(t\right)=\frac{1}{2{\rm{\pi }}}{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}\omega {T}_{\omega }\left(t\right){{\rm{e}}}^{-{\rm{i}}\omega t}{\widetilde{E}}^{{\rm{in}}}\left(\omega \right)=\frac{1}{2\pi }{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}\omega {T}_{\omega }\left(t\right){{\rm{e}}}^{-{\rm{i}}\omega t}{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}{t}^{{\prime} }{{\rm{e}}}^{{\rm{i}}\omega {t}^{{\prime} }}{E}^{\text{in}}\left({t}^{{\prime} }\right).$$
(10)

The transmission of a plane-parallel plane with static properties is known to be

$${T}_{\omega }=\frac{1}{\cos \left({\delta }_{\omega }\right)-\frac{{\rm{i}}}{2}\left({n}_{\omega }+\frac{1}{{n}_{\omega }}\right)\sin \left({\delta }_{\omega }\right)},$$
(11)

where

$${\delta }_{\omega }=\frac{\omega }{c}{n}_{\omega }d,$$

and *d* is the thickness.

For a generalized time-dependent refractive index to describe the transmission through a thin sample, we define ${n}_{\omega }\left(t\right)$ via the time-dependent transmission:

$${T}_{\omega }\left(t\right)=\frac{1}{\cos \left({\delta }_{\omega }\left(t\right)\right)-\frac{{\rm{i}}}{2}\left({n}_{\omega }\left(t\right)+\frac{1}{{n}_{\omega }\left(t\right)}\right)\sin \left({\delta }_{\omega }\left(t\right)\right)},\tag{12}$$

$${\delta }_{\omega }\left(t\right)=\frac{\omega }{c}{n}_{\omega }\left(t\right)d.\tag{13}$$

According to this definition, the effective refractive index may depend on the sample thickness, but this dependence disappears in the limit of an infinitesimally thin sample. To show this, we first consider $\left|{\delta }_{\omega }\left(t\right)\right|\ll 1$ and simplify the above two equations:

$${T}_{\omega }\left(t\right)=1+\frac{{\rm{i}}\omega d}{2c}\left[1+{n}_{\omega }^{2}\left(t\right)\right].\tag{14}$$

Let us now relate ${n}_{\omega }\left(t\right)$ to ${\chi }_{\omega }^{(1)}(t)$, which we defined earlier. Light transmitted through a sample is the superposition of the incident wave and that emitted by the electric current induced in the medium. If light propagates along the *z* axis, while the electric current $J\left(z,t\right)$ is confined to $0\le z\le d$ and flows in a direction perpendicular to the *z* axis, then Maxwell's equations give the following expression for the electric field induced by the electric current:

$${E}^{{\rm{sheet}}}\left(z,t\right)=-\,\frac{2{\rm{\pi }}}{c}{\int }_{0}^{d}{\rm{d}}{z}^{{\prime} }J\left({z}^{{\prime} },t-\frac{\left|z-{z}^{{\prime} }\right|}{c}\right).$$

The field of a transmitted monochromatic wave is then given by

$${E}_{\omega }^{\text{out}}\left(t\right)={e}^{{\rm{i}}\frac{\omega }{c}d-{\rm{i}}\omega t}{E}_{\omega }^{\left(0\right)}-\frac{2\pi }{c}{\int }_{0}^{d}{\rm{d}}{z}^{{\prime} }{J}_{\omega }\left({z}^{{\prime} },t-\frac{\left|d-{z}^{{\prime} }\right|}{c}\right).$$

For an infinitesimally thin sample, the dependence of J on z can be neglected, and we arrive at

$${E}_{\omega }^{\text{out}}\left(t\right)={{\rm{e}}}^{-{\rm{i}}\omega t}\left(1+{\rm{i}}\frac{\omega }{c}d\right){E}_{\omega }^{\left(0\right)}-\frac{2\pi d}{c}{J}_{\omega }\left(t\right).$$
(15)

By comparing this expression with equation (9), we derive

$${T}_{\omega }\left(t\right)=1+\frac{{\rm{i}}\omega d}{c}\left(1+\frac{2{\rm{\pi }}{\rm{i}}}{\omega }\frac{{J}_{\omega }\left(t\right)}{{E}_{\omega }^{\left(0\right)}}{{\rm{e}}}^{{\rm{i}}\omega t}\right).$$

By comparing this expression with equation (14), we relate the time-dependent refractive index to the electric current density:

$${n}_{\omega }^{2}\left(t\right)=1+\frac{4{\rm{\pi }}{\rm{i}}}{\omega }\frac{{J}_{\omega }\left(t\right)}{{E}_{\omega }^{\left(0\right)}}{{\rm{e}}}^{{\rm{i}}\omega t}.$$
(16)

We see that ${n}_{\omega }\left(t\right)$ indeed does not depend on d.

The electric current density ${J}_{\omega }\left(t\right)$ is the time derivative of the polarization induced by a monochromatic wave:

$${J}_{\omega }\left(t\right)={P}_{\omega }^{{\prime} }\left(t\right)=\frac{{\rm{d}}}{{\rm{d}}t}\left({{\rm{e}}}^{-{\rm{i}}\omega t}{\chi }_{\omega }\left(t\right){E}_{\omega }^{\left(0\right)}\right)={{\rm{e}}}^{-{\rm{i}}\omega t}{E}_{\omega }^{\left(0\right)}\left({\chi }_{\omega }^{{\prime} }\left(t\right)-{\rm{i}}\omega {\chi }_{\omega }\left(t\right)\right).$$

Substituting this equation into equation (16), we arrive at the sought after relationship between the generalized refractive index and the generalized linear susceptibility:

$${n}_{\omega }^{2}\left(t\right)=1+4{\rm{\pi }}\left(1+\frac{{\rm{i}}}{\omega }\frac{{\rm{d}}}{{\rm{d}}t}\right)\,{\chi }_{\omega }\left(t\right).$$
(17)

Although we derived this equation for an infinitesimally thin film, one can still use it in equation (12), as a first approximation, for the transmission of a sample with a small but finite thickness.

Drude–Lorentz model for instantaneous photoinjection

The optical properties of conducting solids are well approximated by the Drude–Lorentz model. This is a combination of the Drude model, which describes the motion of free charge carriers, and the Lorentz oscillator model, which describes various resonances that shape the polarization response. The rigorous derivation of the Drude–Lorentz model requires a quantum-mechanical treatment, but essentially the same result can be obtained classically. Here we repeat the the classical derivation for the case in which a solid suddenly changes its properties. At this moment, free charge carriers appear, which are accelerated by the electric field of the test pulse according to Newton's second law:

$${x}^{{\prime\prime} }\left(t\right)+{\gamma }_{\text{D}}{x}^{{\prime} }\left(t\right)=-\,\frac{e}{{m}_{\text{D}}}E\left(t\right).$$
(18)

Here *x* is the electron displacement caused by the external electric field $E\left(t\right)$, subscript 'D' stands for 'Drude', ${\gamma }_{\text{D}}$ is the rate of momentum relaxation, $e > 0$ is the elementary charge and ${m}_{\text{D}}$ is the average effective mass of charge carriers.

In addition to creating charge carriers, photoexcitation enables interband transitions that could not take place in the unperturbed solid: once an electron makes a transition from a valence state to a state in a conduction band, the electron can then be further excited into a higher conduction band by absorbing a photon, whereas the vacancy left in the valence band can be filled by photoexciting an electron from a deeper valence band. Each such transition gives rise to a separate Lorentz term, but it is often sufficient to

consider just a few of them to approximate the linear polarization response in a limited spectral range. In the classical description, the Lorentz contribution to the polarization response emerges from the solution of the following form of Newton's equation for the electron displacement:

$${x}^{{\prime\prime} }\left(t\right)+2{\gamma }_{\text{L}}{x}^{{\prime} }\left(t\right)+{\omega }_{{\rm{r}}}^{2}x\left(t\right)=-\,\frac{e}{{m}_{\text{L}}}E\left(t\right),$$
(19)

where ${\omega }_{\text{r}}$ is the frequency of a Lorentz resonance, γ_{L} is its relaxation rate, and m_{L} is an effective mass. For simplicity, we will consider just one Lorentz term, and we will solve the differential equations with the following initial conditions: $x\left({t}_{0}\right)=0$, ${x}^{{\prime} }\left({t}_{0}\right)=0$, where t_0 is the moment of sudden photoexcitation. The polarization response is related to the classical electron displacement via $P\left(t\right)=-\,{eNx}\left(t\right)$. For the Drude response, N is the concentration of charge carriers; for the Lorentz response, it is the concentration of Lorentz oscillators. To write the final result without explicitly using the concentrations and effective masses, we introduce plasma frequencies. In CGS units, their standard definitions are

$${\omega }_{\text{pl},\text{D}}=\sqrt{\frac{4{\rm{\pi }}{e}^{2}{N}_{{\rm{D}}}}{{m}_{\text{D}}}},$$
$${\omega }_{\text{pl},\text{L}}=\sqrt{\frac{4{\rm{\pi }}{e}^{2}{N}_{{\rm{L}}}}{{m}_{\text{L}}}}.$$

We express the polarization response to a monochromatic wave, ${e}^{-{\rm{i}}\omega t}$, as ${P}_{\omega }\left(t\right)={\chi }_{\omega }\left(t\right){e}^{-{\rm{i}}\omega t}$, solve the above differential equations and obtain the following equations for the time-dependent linear susceptibility in the presence of instantaneous photoinjection:

$${\chi }_{\omega }\left(t\right)={\chi }_{\omega }^{\left(0\right)}+\varDelta {\chi }_{\omega }^{\text{Drude}}\left(t\right)+\varDelta {\chi }_{\omega }^{\text{Lorentz}}\left(t\right),$$
(20)

$$\varDelta {\chi }_{\omega }^{\text{Drude}}\left(t+{t}_{0}\right)=-\,\varTheta \left(t\right)\frac{{\omega }_{\text{pl},\text{D}}^{2}}{4{\rm{\pi }}}\left[\frac{1}{{\omega }^{2}+{\rm{i}}{\gamma }_{\text{D}}\omega }+{\rm{i}}\frac{{{\rm{e}}}^{{\rm{i}}\omega t}}{{\gamma }_{\text{D}}}\left(\frac{1}{\omega }-\frac{{{\rm{e}}}^{-{{\rm{\gamma }}}_{\text{D}}t}}{\omega +{\rm{i}}{\gamma }_{\text{D}}}\right)\right],$$
(21)

and

$$\Delta {\chi }_{\omega }^{\text{Lorentz}}\left(t+{t}_{0}\right)=\varTheta \left(t\right)\frac{{\omega }_{\text{pl},\text{L}}^{2}}{4{\rm{\pi }}}\left[\frac{1}{{\omega }_{{\rm{r}}}^{2}-2{\rm{i}}{\gamma }_{\text{L}}\omega -{\omega }^{2}}-\frac{{{\rm{e}}}^{-{\gamma }_{{\rm{L}}}t}{{\rm{e}}}^{{\rm{i}}\omega t}}{2\varOmega }\left(\frac{{{\rm{e}}}^{-{\rm{i}}\varOmega t}}{\varOmega -\omega -{\rm{i}}{\gamma }_{\text{L}}}+\frac{{{\rm{e}}}^{{\rm{i}}\varOmega t}}{\varOmega +\omega +{\rm{i}}{\gamma }_{\text{L}}}\right)\right],$$
(22)

where

$$\varOmega =\sqrt{{\omega }_{{\rm{r}}}^{2}-{\gamma }_{{\rm{L}}}^{2}}.$$
(23)

In this implementation, the model has five adjustable parameters: ${\omega }_{{\rm{pl}},{\rm{D}}}$, ${\omega }_{{\rm{pl}},{\rm{L}}}$, ${\gamma }_{{\rm{D}}}$, ${\gamma }_{{\rm{L}}}$ and ${\omega }_{{\rm{r}}}$. The susceptibilities of unperturbed samples, ${\chi }_{\omega }^{\left(0\right)}$, were determined from separate field-resolved measurements, where the pump pulse was blocked and test waveforms were recorded with and without the sample. We then used equation (11) to evaluate ${\chi }_{\omega }^{(0)}$ and the sample thickness.

The reconstruction results that we present in the main text were obtained by numerically optimizing the fit parameters, for which we minimized the difference between the measured and reconstructed waveform distortions. The reconstructed waveforms were evaluated with the aid of equations (10), (12), (13), (17) and (20)–(23). Extended Data Table 1 lists the obtained values of the fit parameters.

Nonlinear work

In Fig. 1b, we show the time-dependent nonlinear work (green curve) as a means for illustrating the dynamics of photoinjection. We obtained these data using the SALMON code[28], which implements the real-space real-time time-dependent density functional theory. Using the Tran–Blaha meta-generalized-gradient-approximation exchange potential with Perdew–Wang correlation[41], we calculated the macroscopic electric current density, $J(t)$, induced by a realistic laser pulse shown in Extended Data Fig. 1. The peak electric field of the pulse inside the silicon crystal was set to 0.35 V Å^{-1}. The time-dependent work performed by the electric field $E(t)$ is given by $W\left(t\right)={\int }_{-\infty }^{t}E\left({t}^{{\prime} }\right)J\left({t}^{{\prime} }\right){{\rm{d}}t}^{{\prime} }$. To get the nonlinear work, we ran another simulation with a much weaker pulse (0.001 V Å^{-1}), which gave us a good estimation of the linear work. In the absence of single-photon transitions, the linear work is responsible for the linear polarization that the pulse transiently induces in the medium. The linear work scales as the square of the peak electric field, which allows one to calculate the linear component of the work performed by an intense pulse. By subtracting the linear work from the total work, we obtained the nonlinear component of the total work, which is due to nonlinear interband transitions and, to a lesser extent, the motion of photoinjected charge carriers.

Optical system

A Ti:sapphire oscillator (Rainbow 2, Spectra Physics) was used to provide an octave-spanning bandwidth, centred at 750 nm (Extended Data Fig. 2). The carrier-envelope-phase (CEP) from the oscillator was stabilized using a feed-forward scheme. The pulses were further amplified within a nine-pass

cryo-cooled Ti:sapphire chirped-pulse amplifier at a repetition rate of 3 kHz and temporally compressed using a transmission grating-based compressor, yielding 21 fs pulses with 2.5 W output power. The amplified pulses were spectrally broadened in a hollow-core fibre filled with neon gas (1.8 bar pressure), resulting in a spectral broadening that covered the wavelength range from 400 nm to 1,100 nm. Three pairs of chirped mirrors in combination with 6 mm of fused silica (and 50 cm of air) were used to compress the pulses to about 3 fs (full-width half-maximum of the intensity envelope, evaluated from nonlinear photoconductive sampling[3] measurements). The compressed pulses were then guided to the experimental setup for further experiments. An extra CEP stabilization loop, for long-term drifts, was implemented using the fibre output and the f-to-$2f$ technique.

Data acquisition

The electric-field waveforms were recorded with a recently demonstrated GHOST[29] with a heterodyne signal produced in a z-cut α-quartz crystal of about 12.3 µm thickness. The optical signal for GHOST detection was measured with a fast photodiode (Roither Lasertechnik). A transimpedance amplifier (DLPCA-200, FEMTO Messtechnik) was used to provide an amplified voltage signal to a gated integrator (SR250 Stanford Research Systems) that was triggered by an electrical signal synchronized to the repetition rate of the laser. The integrator produced a d.c. voltage that was proportional to the integrated photodiode signal for all combinations of optical pulses (see the next section). The signal from the integrator was recorded using an oscilloscope (Tektronix) for analogue-to-digital conversion and recorded using a computer interface to the General Purpose Interface Bus (National Instruments).

Multiplexed detection of perturbed and unperturbed waveforms

Fluctuations and drifts in the laser system modify the test pulse during a delay scan, which presents a major obstacle for measuring small changes in electric-field waveforms. To mitigate this problem, we measured the modified and reference waveforms almost simultaneously. For each arrival

time of the sampling pulse, we measured the following three signals: (1) a signal generated by the sampling pulse interacting with the modified waveform; (2) a signal for which the sampling pulse interacted with the reference waveform and (3) a signal for which only the sampling pulse reached the photodetector. This form of detection was implemented using optical chopper wheels to transmit every third pump and block every third test pulse, generating the following sequence of pulses:

Pump pulse: 1 0 0

Test pulse: 1 1 0

Sampling pulse: 1 1 1

Here ‘1’ indicates that the pulse is present, whereas ‘0’ indicates that the pulse is blocked by the chopper wheel.

This sequence of pulses is delivered to the photodiode. The current signal from the photodiode is then further amplified and converted to a voltage by means of a transimpedance amplifier (DLPCA-200, FEMTO Messtechnik). The output from a transimpedance amplifier provides a time-domain train of voltage signals (Extended Data Fig. 3a). A boxcar integrator (SR200, Stanford Research Systems) provides a d.c. output for each of these voltage pulses (Extended Data Fig. 3b): ${{\rm{d.c.}}}_{\alpha }$, ${{\rm{d.c.}}}_{\beta }$ and ${\rm{d.}}{{\rm{c.}}}_{\gamma }$ correspond to the three columns in the above table, respectively. The output from the boxcar integrator is digitized and stored. The pump pulse is blocked after the sample and does not reach the detector.

If the pump pulse does not change any properties of the sample medium, the ${{\rm{d.c.}}}_{\alpha }$ and ${{\rm{d.c.}}}_{\beta }$ outputs are identical. If, however, the pump pulse changes the sample medium, then the difference between ${{\rm{d.c.}}}_{\alpha }$ and ${{\rm{d.c.}}}_{\beta }$ outputs corresponds to a difference in the test waveform due to the

change of the sample properties. The signal related to the modified waveform is ${\rm{d.}}{{\rm{c.}}}_{\alpha }-{\rm{d.}}{{\rm{c.}}}_{\gamma }$, whereas the signal related to the reference waveform is ${\rm{d.}}{{\rm{c.}}}_{\beta }-{\rm{d.}}{{\rm{c.}}}_{\gamma }$.

The experimental setup

The optical setup is shown in Extended Data Fig. 4. The pump pulse, obtained from the reflection off the front surface of the first wedge in the wedge pair (WP 1), was delayed using a closed-loop piezo stage (PX 200, Piezosystems Jena). The sampling pulse is obtained by taking the reflection from a 200-mm-thick ultraviolet-grade fused silica window (FS 2). The reflected light was delayed using another closed-loop piezo stage (PX 200, Piezosystems Jena). Wire-grid polarizers (WG 1, WG 2, WG 3 and WG 5) were used to control the energies of the test, pump and sampling pulses, whereas fused-silica wedge pairs (WP 1, WP 2, WP 3) were used for fine dispersion compensation and CEP control. An off-axis protected silver parabolic mirror (OPM 2) was used to focus the test arm on a 0.8-mm-thick ($\theta ={10.3}^{^\circ }$, where θ is the angle between the surface normal and the crystal axis) BiBo crystal (NL 1) for intra-pulse difference-frequency generation. A set of fused-silica and silicon plates (FS 1, Si) was used for the compression of the near-infrared pulses and for blocking the fundamental, respectively. The near-infrared light was collimated by a protected gold off-axis parabolic mirror (OPM 3).

The pump and test pulses were recombined using a wire-grid polarizer (WG 4), which transmitted the pump pulses and reflected the test pulses. The orthogonally polarized pulses were focused using a protected silver off-axis parabolic mirror (OPM 1) on the sample and re-collimated using another protected silver off-axis parabolic mirror (OPM 4). The pump pulses were transmitted through a hole in the re-collimation mirror (OPM 4). The collimated beam was recombined with the sampling pulses using a wire-grid polarizer (WG 6). A 12.34-mm-thick *z*-cut quartz crystal (NL 2) upconverted the sampling pulse. After spectral filtering by a bandpass filter (BP 1) and polarization control by a wire-grid polarizer (WG 7), this high-frequency light served as the local oscillator for generalized heterodyne

optical sampling. At the same time, NL 2 enabled nonlinear wave mixing between the sampling and test pulses. The interference of this signal with the local oscillator was detected with a SiC-based photodiode (PD 1). Modulation of the test and pump pulses was accomplished using chopper wheels (CW 1 and CW 2), generating the sequence of pulses for the almost simultaneous waveform detection.

Timing of the injection event

The following experimental procedure was used to determine the moment of photoinjection (delay zero). The experimental setup has two focal planes: the sampling and detection ones. The pump and test pulses interact with a thin sample in the sample plane. Before the transmitted test pulse reaches the detection plane, the transmitted pump pulse is removed and a sampling pulse is added to the beam. GHOST detection then takes place in the detection plane, where the two pulses interact with the nonlinear crystal that is labelled as NL 2 in Extended Data Fig. 4. The wedge pair WP 2 was used to set the CEP of the pump pulse in the sample plane to 0 (cosine pulse). The wedge pair WP 3 was used to set the CEP of the sampling pulse in the detection plane to 0. The pump pulse in the sample plane is a replica of the sampling pulse in the detection plane because both pulses are derived from the same incoming pulse, they have the same CEP and they propagate through the wedges made of the same medium under conditions adjusted for maximum compression.

In the detection plane, the strong sampling pulse produces a nonlinear gate confined to the vicinity of the strongest peak of the sampling pulse. The gate was scanned through the test pulse by a retroreflector consisting of mirrors M6 and M7 placed on piezo translation stage TS 2. As the nonlinear gate is confined to the vicinity of the strongest peak of the sampling pulse, the signal produced at each position of the TS 2 corresponds to the temporal overlap between the strongest peak of the sampling pulse and the part of the test pulse within the window of the duration of the gate (less than 1 fs). Hence, by sampling the test-pulse waveform, one can the map positions of translation stage TS 2 to temporal overlaps of the peak of the sampling pulse with various parts of the test waveform.

To determine the moment of photoinjection in the sample plane (delay zero with respect to the test pulse), the sample was removed and a nonlinear crystal NL 2 was installed in the sample plane. Then the test pulse was measured in the sample plane using the same GHOST scheme as in the detection plane but with the pump pulse playing the role of the sampling pulse. This not only allows one to characterize the test pulse in the sample plane but also to map the positions of the translation stage TS 1 to temporal overlaps of the peak of the pump pulse with various parts of the test pulse. Extended Data Fig. 5 shows an example of such a pair of measurements.

After measuring the test-pulse waveform in both sampling and detection planes, the relation between the positions of translation stages TS 1 and TS 2 can be determined. Hence, various moments of photoinjection in the sample plane can be projected onto the test-pulse waveform measured in the detection plane. When the injection translation stage is moved to position x, this corresponds to the excitation of the sample at the location on the test-pulse waveform that corresponds to the position of the translation stage in the detection plane being $y=x+d$.

The constant shift d was measured before each experimental campaign by sampling the test pulse in the sample and detection planes with the same optics and electronics. Once d is determined, the moment of the photoinjection can be mapped directly on the measured test-pulse waveform in the detection plane.

Data availability

The measured and simulated data for all the figures in this paper are available from the Edmond Data Repository of the Max Planck Society[42]. Source data are provided with this paper.

Code availability

The codes used to reconstruct the measured waveform distortions and prepare the plots are available from the Edmond Data Repository of the Max Planck Society[42].

References

1. Auston, D. H., Cheung, K. P. & Smith, P. R. Picosecond photoconducting Hertzian dipoles. *Appl. Phys. Lett.* **45**, 284–286 (1984).

2. Ashida, M. et al. Ultrabroadband THz field detection beyond 170THz with a photoconductive antenna. In *Proc. 2008 Conference on Lasers and Electro-Optics & Quantum Electronics and Laser Science* 1–2 (IEEE, 2008).

3. Sederberg, S. et al. Attosecond optoelectronic field measurement in solids. *Nat. Commun.* **11**, 430 (2020).

4. Zimin, D. et al. Petahertz-scale nonlinear photoconductive sampling in air. *Optica* **8**, 586–590 (2021).

5. Bionta, M. R. et al. On-chip sampling of optical fields with attosecond resolution. *Nat. Photonics* **15**, 456–460 (2021).

6. Liu, Y. Y. et al. All-optical sampling of few-cycle infrared pulses using tunneling in a solid. *Photon. Res.* **9**, 929–936 (2021).

7. Hui, D. et al. Attosecond electron motion control in dielectric. *Nat. Photonics* **16**, 33–37 (2022).

8. Huber, R. et al. How many-particle interactions develop after ultrafast excitation of an electron-hole plasma. *Nature* **414**, 286–289 (2001).

9. Leitenstorfer, A. et al. Femtosecond buildup of Coulomb screening in a photoexcited electron-hole plasma. *Physica B* **314**, 248–254 (2002).

10. Hase, M., Kitajima, M., Constantinescu, A. M. & Petek, H. The birth of a quasiparticle in silicon observed in time-frequency space. *Nature* **426**, 51–54 (2003).

11. Kira, M. & Koch, S. W. Many-body correlations and excitonic effects in semiconductor spectroscopy. *Prog. Quantum Electron.* **30**, 155–296

(2006).

12. Sangalli, D., Dal Conte, S., Manzoni, C., Cerullo, G. & Marini, A. Nonequilibrium optical properties in semiconductors from first principles: a combined theoretical and experimental study of bulk silicon. *Phys. Rev. B* https://doi.org/10.1103/PhysRevB.93.195205 (2016).

13. Gaal, P. et al. Internal motions of a quasiparticle governing its ultrafast nonlinear response. *Nature* **450**, 1210–1213 (2007).

14. Goulielmakis, E. et al. Single-cycle nonlinear optics. *Science* **320**, 1614–1617 (2008).

15. Schultze, M. et al. Controlling dielectrics with the electric field of light. *Nature* **493**, 75–78 (2013).

16. Sommer, A. et al. Attosecond nonlinear polarization and light-matter energy transfer in solids. *Nature* **534**, 86–90 (2016).

17. Paasch-Colberg, T. et al. Sub-cycle optical control of current in a semiconductor: from the multiphoton to the tunneling regime. *Optica* **3**, 1358–1361 (2016).

18. Schiffrin, A. et al. Optical-field-induced current in dielectrics. *Nature* **493**, 70–74 (2013).

19. Krausz, F. & Stockman, M. I. Attosecond metrology: from electron capture to future signal processing. *Nat. Photonics* **8**, 205–213 (2014).

20. Mayer, B. et al. Sub-cycle slicing of phase-locked and intense mid-infrared transients. *New J. Phys.* https://doi.org/10.1088/1367-2630/16/6/063033 (2014).

21. Mashiko, H., Oguri, K., Yamaguchi, T., Suda, A. & Gotoh, H. Petahertz optical drive with wide-bandgap semiconductor. *Nat. Phys.* **12**, 741–745 (2016).

22. Kruchinin, S. Y., Krausz, F. & Yakovlev, V. S. Colloquium: strong-field phenomena in periodic systems. *Rev. Mod. Phys.* **90**, 21002 (2018).

23. Schoetz, J. et al. Perspective on petahertz electronics and attosecond nanoscopy. *ACS Photonics* **6**, 3057–3069 (2019).

24. El Sayed, K., Schuster, S., Haug, H., Herzel, F. & Henneberger, K. Subpicosecond plasmon response: buildup of screening. *Phys. Rev. B* **49**, 7337–7344 (1994).

25. Bányai, L., Vu, Q. T., Mieck, B. & Haug, H. Ultrafast quantum kinetics of time-dependent RPA-screened Coulomb scattering. *Phys. Rev. Lett.* **81**, 882–885 (1998).

26. Borisov, A., Sanchez-Portal, D., Muino, R. D. & Echenique, P. M. Building up the screening below the femtosecond scale. *Chem. Phys. Lett.* **387**, 95–100 (2004).

27. Hugel, W. A. et al. Photon echoes from semiconductor band-to-band continuum transitions in the regime of Coulomb quantum kinetics. *Phys. Rev. Lett.* **83**, 3313–3316 (1999).

28. Noda, M. et al. SALMON: Scalable Ab-initio Light-Matter simulator for Optics and Nanoscience. *Comput. Phys. Commun.* **235**, 356–365 (2019).

29. Zimin, D. A., Yakovlev, V. S. & Karpowicz, N. Ultra-broadband all-optical sampling of optical waveforms. *Sci. Adv.* **8**, eade1029 (2022).

30. Boyd, R. W. *Nonlinear Optic*s 4th edn (Academic Press, 2019).

31. Nemec, H., Kadlec, F., Surendran, S., Kuzel, P. & Jungwirth, P. Ultrafast far-infrared dynamics probed by terahertz pulses: a frequency domain approach. I. Model systems. *J. Chem. Phys.* **122**, 104503 (2005).

32. Schins, J. M., Hendry, E., Bonn, M. & Muller, H. G. Retrieving the susceptibility from time-resolved terahertz experiments. *J. Chem. Phys.* **127**, 094308 (2007).

33. Orenstein, J. & Dodge, J. S. Terahertz time-domain spectroscopy of transient metallic and superconducting states. *Phys. Rev. B* **92**, 134507 (2015).

34. Yacoby, Y. High-frequency Franz-Keldysh effect. *Phys. Rev.* **169**, 610 (1968).

35. Jauho, A. P. & Johnsen, K. Dynamical Franz-Keldysh effect. *Phys. Rev. Lett.* **76**, 4576–4579 (1996).

36. Yamada, S. & Yabana, K. Determining the optimum thickness for high harmonic generation from nanoscale thin films: an ab initio computational study. *Phys. Rev. B* https://doi.org/10.1103/PhysRevB.103.155426 (2021).

37. van Schilfgaarde, M., Kotani, T. & Faleev, S. Quasiparticle self-consistent GW theory. *Phys. Rev. Lett.* **96**, 226402 (2006).

38. Stefanucci, G. & Leeuwen, R. v. *Nonequilibrium Many-body Theory of Quantum Systems: A Modern Introduction* (Cambridge University Press, 2013).

39. Marques, M. A. L., Maitra, N. T., Nogueira, F. M. S., Gross, E. K. U. & Rubio, A. *Fundamentals of Time-Dependent Density Functional Theory* (Springer, 2012).

40. Haug, H. & Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductors* 4th edn (World Scientific, 2009).

41. Tran, F. & Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102**, 226401.

42. Dynamic optical response of solids following 1-fs-scale photoinjection. *Edmond Data Repository* https://doi.org/10.17617/3.VKRMLS (2023).

Acknowledgements

We thank S. Sederberg and M. Uemoto for discussions. This research is based upon work supported by the Munich Centre for Advanced Photonics. N.K. was partially supported by the FISR-CNR project 'TECNOMED—Tecnopolo di nanotecnologia e fotonica per la medicina di precision'. M.Q. and D.A.Z. acknowledge support from the Max Planck Society via the International Max Planck Research School for Advanced Photon Science (IMPRS-APS).

Funding

Open access funding provided by the Max Planck Society.

Author information

Authors and Affiliations

1. Max-Planck-Institut für Quantenoptik (MPQ), Garching, Germany

   Dmitry A. Zimin, Nicholas Karpowicz, Muhammad Qasim, Matthew Weidman, Ferenc Krausz & Vladislav S. Yakovlev

2. Ludwig-Maximilians-Universität München (LMU), Garching, Germany

   Dmitry A. Zimin, Muhammad Qasim, Matthew Weidman, Ferenc Krausz & Vladislav S. Yakovlev

3. CNR NANOTEC Institute of Nanotechnology, Lecce, Italy

   Nicholas Karpowicz

Contributions

F.K. and N.K. conceived the project. D.A.Z. performed the experiments. D.A.Z., M.Q. and V.S.Y. analysed the experimental data and performed the theoretical modelling and calculations. All authors discussed the results and contributed to writing the manuscript.

Corresponding authors

Correspondence to Nicholas Karpowicz or Vladislav S. Yakovlev.

Ethics declarations

Competing interests

The athors declare no competing interests.

Peer review

Peer review information

Nature thanks Katsuya Oguri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 The electric field of the pump pulse that was used in the TDDFT simulations.

Figure 1b in the main text shows the central part of this waveform, which was obtained by first measuring the pump pulse with the aid of nonlinear photoconductive sampling and then adapting these measurement results for numerical simulations: suppressing the wings and removing the noise by spectral filtering.

Extended Data Fig. 2 Diagram of the laser source layout.

The output of the Ti:Sa oscillator is guided through the CEP4 module (Femtolasers) for CEP stabilization (Stage I carrier-envelope phase stabilization). The transmitted pulse is amplified in a chirped-pulse amplifier (Ti:CPA). The amplified pulse is guided into the hollow core fiber filled with neon gas. The beam is coupled into the fiber following active stabilization. After the hollow core fiber, the beam passes through three pairs of chirped mirrors (Ultrafast Innovations) which compensate for the spectral phase of the incident pulse as well as the additional dispersion of the experimental setup. In front of the compressor, an outcoupling mirror reflects a small fraction of the beam into the second stage of the carrier-envelope phase stabilization, which is based on the f-2f scheme.

Extended Data Fig. 3 Schematic representation of the multiplexing scheme.

a, A photodiode registers three consecutive optical pulses labeled as α, β, and γ. Each pulse produces an electrical signal with the shape of a decaying exponential function (blue). The integral of this signal is proportional to the energy carried by the respective optical pulse. The orange outlines represent the windows of the integration time. **b**, The boxcar integrator generates an output d.c. signal after each integration window. This d.c. value can be further read out with an oscilloscope or any analog-to-digital converter device.

Extended Data Fig. 4 Schematic of the experimental setup.

An optical pulse from the laser source is split into three optical arms: test, injection and sampling. The pulse in the test arm is down-converted to a pulse with a central wavelength of 2.1 μm. The test and injection pulses are combined with a controlled delay between them and transmitted through a sample. The injection pulse is then filtered out, while the transmitted test pulse is combined with the sampling pulse, which nonlinearly interacts with the test pulse in the detection plane, sampling its electric field. This figure was created with the online 3Doptix platform (https://3doptix.com/).

Extended Data Fig. 5 Timing (translation stage positions) synchronization between two focal planes.

Since photoinjection takes place in the sample plane, while the test pulse is sampled in the detection plane, it is necessary to calibrate the delay between the pump and test pulses. This is done by determining the mapping between the positions of piezo stages that control the pump-test and sampling-test delays, which is accomplished by comparing test waveforms measured in the sample and detection planes.

Extended Data Table 1 The retrieved values of the fit parameters

Supplementary information

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Zimin, D.A., Karpowicz, N., Qasim, M. *et al.* Dynamic optical response of solids following 1-fs-scale photoinjection. *Nature* **618**, 276–280 (2023). https://doi.org/10.1038/s41586-023-05986-w

- Received: 22 December 2021
- Accepted: 21 March 2023
- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05986-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

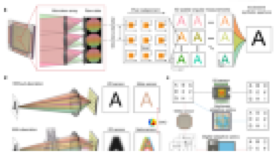
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

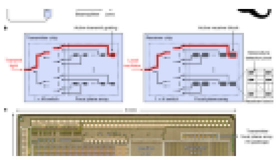
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05986-w



An integrated imaging sensor for aberration-corrected 3D photography

19 October 2022

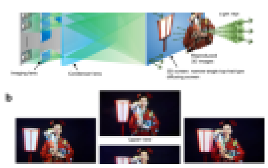
Jiamin Wu, Yuduo Guo, … Qionghai Dai



A universal 3D imaging sensor on a silicon photonics platform

10 February 2021

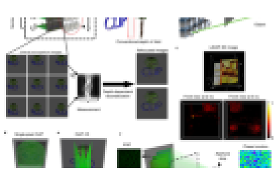
Christopher Rogers, Alexander Y. Piggott, … Remus Nicolaescu



Aktina Vision: Full-parallax three-dimensional display with 100 million light rays

27 November 2019

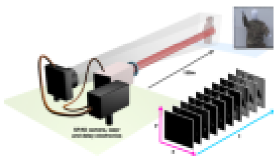
Hayato Watanabe, Naoto Okaichi, … Masahiro Kawakita



[Compact light field photography towards versatile three-dimensional vision]

09 June 2022

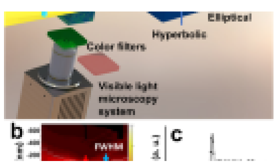
Xiaohua Feng, Yayao Ma & Liang Gao



[Long-range depth imaging using a single-photon detector array and non-local data fusion]

30 May 2019

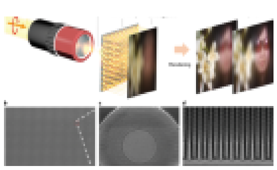
Susan Chan, Abderrahim Halimi, … Jonathan Leach



[The new X-ray/visible microscopy MAXWELL technique for fast three-dimensional nanoimaging with isotropic resolution]

11 June 2022

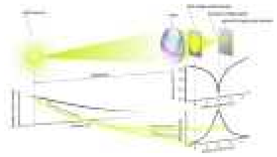
Yoshiki Kohmura, Shun-Min Yang, … Yeukuang Hwu



[Achromatic metalens array for full-colour light-field imaging]

21 January 2019

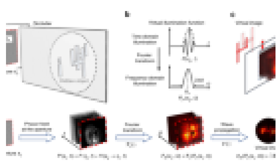
Ren Jie Lin, Vin-Cent Su, … Din Ping Tsai



[**Focus-Induced Photoresponse: a novel way to measure distances with photodetectors**]

15 June 2018

Oili Pekkola, Christoph Lungenschmied, … Ingmar Bruder



[**Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems**]

02 April 2020

Xiaochun Liu, Sebastian Bauer & Andreas Velten

- Article
- Open Access
- Published: 10 May 2023

X-ray-to-visible light-field detection through pixelated colour conversion

- Luying Yi,
- Bo Hou,
- He Zhao &

- …
- Xiaogang Liu

Show authors

Nature **volume 618**, pages 281–286 (2023)

- 9719 Accesses
- 99 Altmetric
- Metrics details

Abstract

Light-field detection measures both the intensity of light rays and their precise direction in free space. However, current light-field detection techniques either require complex microlens arrays or are limited to the ultraviolet–visible light wavelength ranges[1,2,3,4]. Here we present a robust, scalable method based on lithographically patterned perovskite nanocrystal arrays that can be used to determine radiation vectors from X-rays to visible light (0.002–550 nm). With these multicolour nanocrystal arrays, light rays from specific directions can be converted into pixelated colour outputs with an angular resolution of 0.0018°. We find that three-dimensional light-field detection and spatial positioning of light sources are possible by modifying nanocrystal arrays with specific orientations. We also demonstrate three-dimensional object imaging and visible light and X-ray phase-contrast imaging by combining pixelated nanocrystal arrays with a colour charge-coupled device. The ability to detect light direction beyond optical wavelengths through colour-contrast encoding could enable new applications, for example, in three-dimensional phase-contrast imaging, robotics, virtual reality, tomographic biological imaging and satellite autonomous navigation.

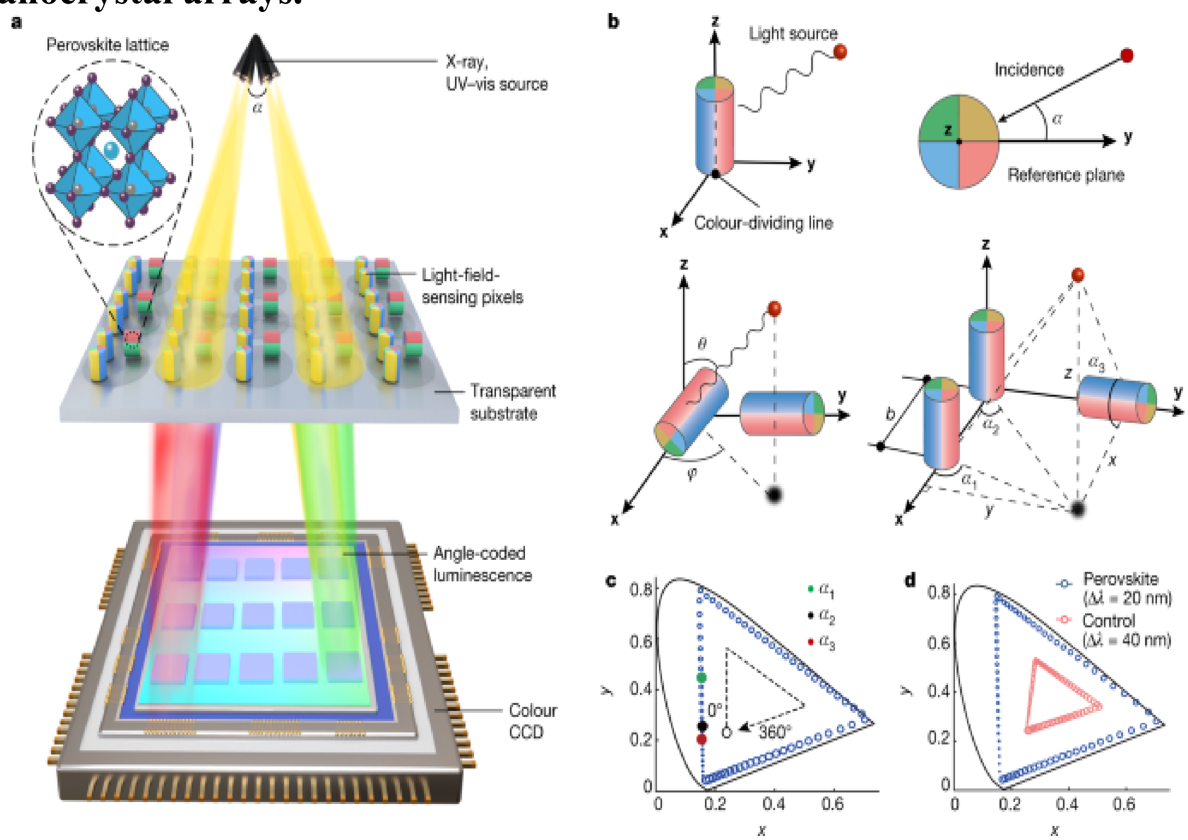
Main

Advances in materials and semiconductor processes have revolutionized the design and fabrication of micro- and nano-photodetectors. But the pixels of most sensors detect only the intensity of electromagnetic waves. As a result, all phase information of the objects and diffracted light waves is lost[5,6,7,8,9,10]. Although intensity information alone is sufficient for conventional applications such as two-dimensional photography and microscopy imaging, this limitation hinders three-dimensional (3D) and four-dimensional imaging applications, including phase-contrast imaging, light detection and ranging, autonomous vehicles, virtual reality and space exploration[11,12,13,14,15,16,17,18,19]. An optical array of microlenses or photonic crystals with pixelated photodiodes is usually used to measure the light field or the distribution of light directions and thus to characterize phase information. Nevertheless, integration of these elements into complementary metal-oxide-semiconductor architectures is costly and complex[4,20,21,22]. Optical resonances in subwavelength semiconductor structures enable the development of angle-sensitive structures by manipulating light–matter interactions[23,24,25,26,27,28]. However, most of them depend on wavelength or polarization and require materials with a high refractive index[29]. Moreover, light-vector detection and control are at present limited to ultraviolet- and visible-light wavelengths. Although several sensors using Shack–Hartmann or Hartmann structures are capable of phase measurements in the extreme ultraviolet-light range, phase measurements of hard X-rays and gamma rays remain challenging because high-energy beams cannot be focused using conventional mirrors or microlenses[30,31].

Owing to the versatility of colour encoding in data visualization, we proposed that colour-contrast encoding could be used to visualize the directions of light rays. To test our hypothesis, we selected inorganic perovskite nanocrystals because of their excellent optoelectronic properties[32,33,34,35]. They also demonstrate highly efficient and tunable emission with high colour saturation across the visible spectrum under X-ray or visible-light irradiation. Furthermore, Sn-based perovskite nanocrystals can have optical bandgaps that extend into the near-infrared light region[36,37]. A fundamental design for a 3D light-field detection involves lithographically patterning perovskite nanocrystals onto a transparent substrate (Fig. 1a). A 3D light-field sensor can then be constructed by integrating the patterned thin-film substrate with a colour charge-coupled device (CCD) that converts

the angle of incident light rays into a specific colour output. The basic unit of 3D light-field sensor is a single azimuth detector comprising multicolour-emitting perovskite nanocrystals. When incident light strikes patterned nanocrystals, the azimuth angle α between the incident light and the reference plane can be detected by measuring the colour output of the basic unit (Fig. 1b). Specifically, two azimuth detectors arranged perpendicular to each other can realize 3D light-direction sensing and determine the azimuth angle φ and elevation angle θ of the incident light in spherical coordinates. To determine the absolute position of a light source, three azimuth detectors can be arranged to create a correlation among the three corresponding azimuth angles α_1, α_2 and α_3 encoded in the colour outputs.

Fig. 1: X-ray-to-visible light-field detection using pixelated perovskite nanocrystal arrays.



a, Design of the 3D light-field sensor on the basis of pixelated colour conversion. Light-field-sensing pixels, which consist of patterned perovskite nanocrystals on a transparent film, convert light from different directions into luminescence signals of different colours, which can be detected by a

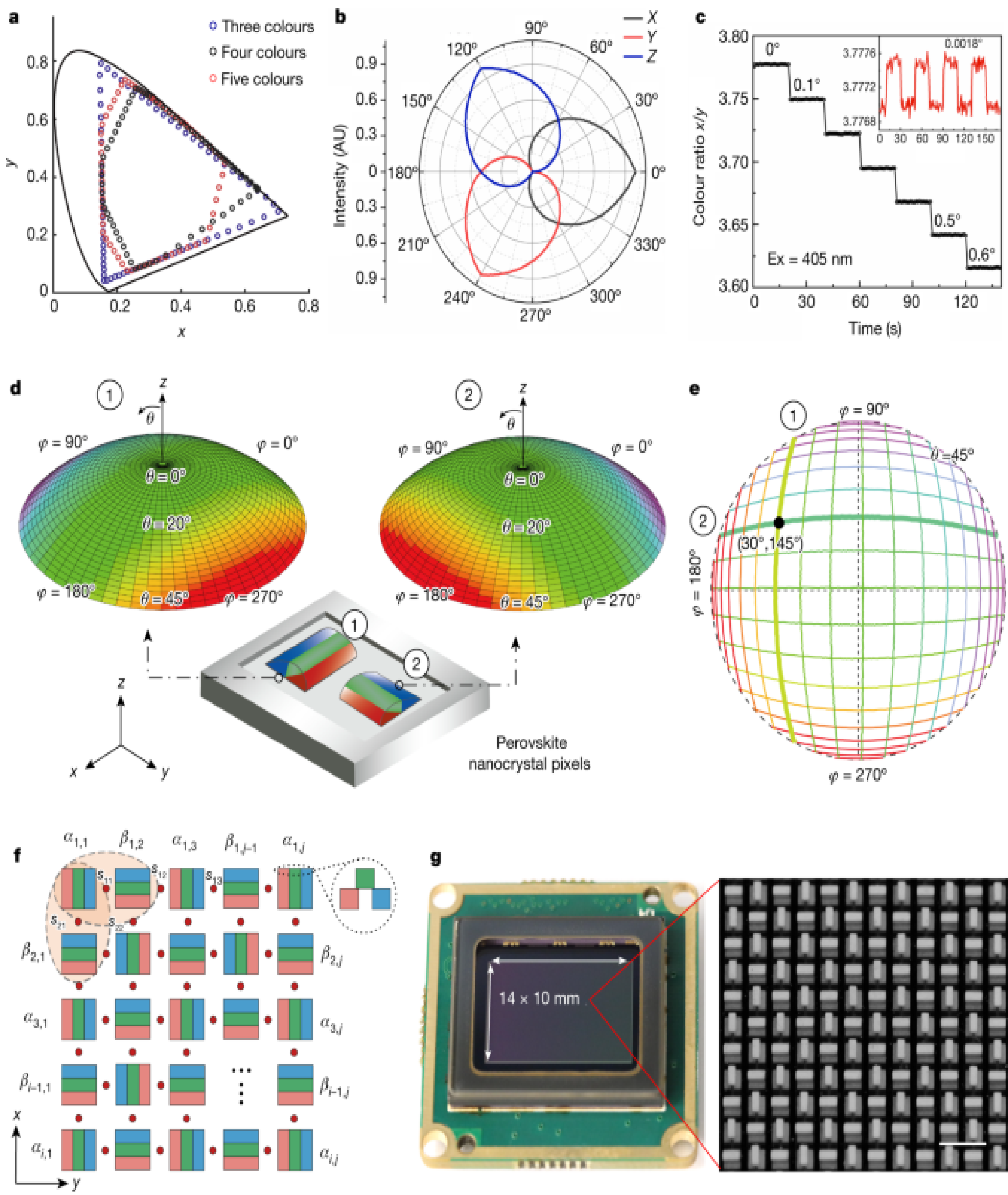
colour CCD. UV–vis, ultraviolet–visible light. **b**, The working principle of light-field sensing by colour conversion. The basic unit of the 3D light-field sensor is a single azimuth detector comprising multicolour-emitting perovskite nanocrystals. The colour of output luminescence depends on the angle α between the incident light and the reference plane. Two perpendicularly arranged azimuth detectors can realize 3D light-field sensing and determine the azimuth angle φ and elevation angle θ of the incident light in spherical coordinates. By arranging three azimuth detectors, correlation of the three azimuth angles α_1, α_2 and α_3 encoded in the colour outputs of the three azimuth detectors enables detection of the absolute position (x, y, z) of a light source. **c**, Chromaticity responses of a single azimuth detector at light incidence from 0° to 360° relative to the reference plane. Red, green and black dots correspond to the three azimuth angles α_1, α_2 and α_3, respectively, recorded using the three azimuth detectors shown in **b**. **d**, Chromaticity response of a single perovskite nanocrystal-based azimuth detector at light incidence from 0° to 360°, relative to the control comprising $ZnS:Cu^{2+}/Mn^{2+}$ and $SrAl_2O_4:Eu^{2+}/Dy^{3+}$ phosphors.

As a proof of concept, we synthesized inorganic perovskite nanocrystals ($CsPbX_3$; X = Cl, Br or I) according to the literature[36,38,39] (Supplementary Information section 1). We selected three sets of perovskite quantum dots with emissions at 445 nm, 523 nm and 652 nm to construct a single azimuth detector. When light is incident from 0° to 360° relative to the reference direction, the detected colour gamut forms a large triangle on the Commission Internationale de l’éclairage (CIE) *xy* chromaticity diagram (Fig. 1c). The position of the colour output on the chromaticity diagram determines the incident angle of the light and a larger triangle indicates a higher angular resolution. We found that the colour gamut of azimuth detectors made of perovskite nanocrystals forms a larger triangle in the chromaticity diagram compared with detectors made of $ZnS:Cu^{2+}/Mn^{2+}$ and $SrAl_2O_4:Eu^{2+}/Dy^{3+}$ phosphors (Fig. 1d). Azimuth detectors generate a higher angular resolution, because of the broader colour coverage and higher colour saturation of perovskite nanocrystals.

Single azimuth detectors with different colour gamuts produce colour plots of varying shapes (Fig. 2a). Nanocrystals with red, green and blue outputs

can detect extremely small angular changes. We used this property and built a single three-colour azimuth detector on a red, green and blue sensor chip that converts incident light from 0° to 360° into different CIE *XYZ* tristimulus values of luminescence (Fig. 2b). The minimum detectable angular change is determined by the contrast ratio of the colour response and by the signal-to-noise ratio (SNR) of the colour sensor. Each primary colour in our tests contains 65,536 levels, which yields a detection limit of around 0.0018° angular change at a wavelength of 405 nm and a power of 8 mW (Fig. 2c).

Fig. 2: Characterization of the pixelated colour conversion for 3D light-field sensing.



a, Chromaticity responses of single azimuthal detectors composed of three-, four- and five-colour perovskite nanocrystals as a function of the direction of incident light. **b**, CIE tristimulus values *X*, *Y* and *Z* of the output luminescence of a single azimuth detector as a function of the direction of incident light. AU, arbitrary units. **c**, Azimuth resolution measurement for visible light (405 nm) using a single azimuth detector, with a minimum detectable angular change of 0.0018°. Excitation (Ex) wavelength, 450 nm.

d, Two types of colour map recorded from two perpendicularly aligned azimuth detectors with light incident from different azimuth angles φ and elevation angles θ. **e**, Contour lines extracted from the two colour maps in **d**. A unique incidence direction can be determined by combining the colour values from two azimuth detectors. **f**, Top view of the azimuthal detector arrays for imaging the 3D light direction, in which adjacent pixels of perovskite nanocrystals are aligned perpendicularly. The two detectors encircled by the ellipse can determine the angle of the beam incident on the centre point of the ellipse. Inset, side view of a patterned pixel. **g**, Photograph of a 3D light-field sensor fabricated by integrating the perovskite nanocrystal array into a colour CCD. Inset, a section of the microscopy image of nanocrystal-based azimuth detectors. Scale bar, 150 μm (**g**).

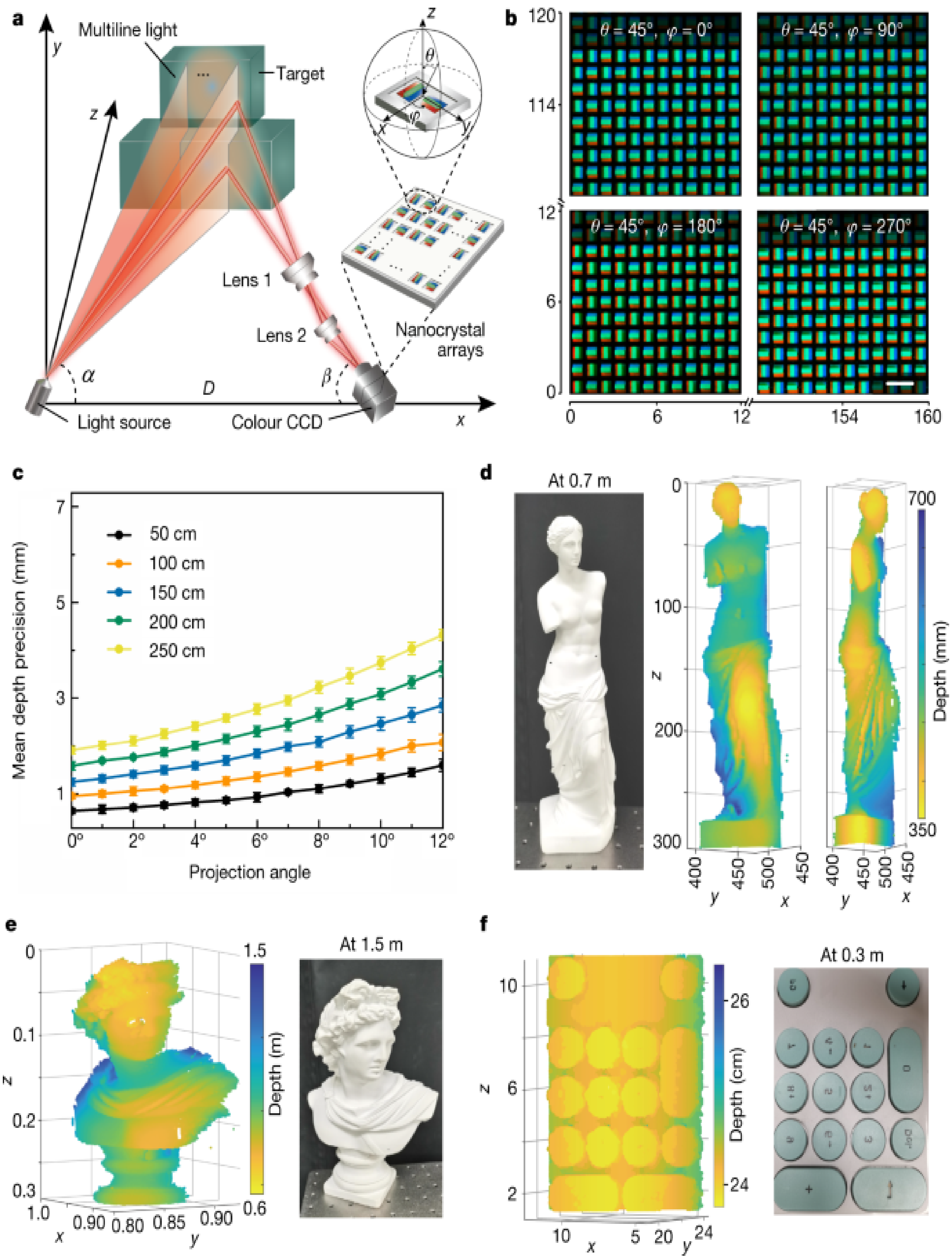
We next designed and fabricated two azimuth detectors arranged perpendicular to each other for omnidirectional light-field detection (Fig. 2d). In spherical coordinates, the azimuth angle φ and elevation angle θ for each incident beam can be calculated using the formula $\varphi = \arctan(\tan \alpha_1/\tan \alpha_2)$ and $\theta = \arctan[(\tan^2 \alpha_1 + \tan^2 \alpha_2)^{1/2}]$, respectively, where α_1 and α_2 are obtained from the emission colours of the two azimuth detectors (Supplementary Information sections 2–4). The two azimuth detectors yielded two types of colour map at different angles of incidence. The contours of the two colour maps in the polar plot enable specific incidence angles to be determined by combining the colour values of two azimuth detectors (Fig. 2e). We further designed azimuthal detector arrays to image the 3D light direction, in which adjacent pixels of perovskite nanocrystals were aligned perpendicular to each other (Fig. 2f). For simplicity, the angle detected by detectors parallel to the x axis is denoted by $\alpha_{i,j}$ (where i and j refer to the rows and columns of the nanocrystal arrays); the angle detected by detectors parallel to the y axis is denoted by $\beta_{i,j}$. Each of the two azimuth detectors, which are perpendicular to each other, can reconstruct the angle of the beam incident at the centre of the two pixels. For example, $\alpha_{1,1}$ and $\beta_{1,2}$ can be used to calculate the 3D angle of the beam incident at point s_{11}, whereas $\beta_{2,1}$ and $\alpha_{1,1}$ can be used to calculate the 3D angle of the beam incident at point s_{21}. Therefore, the imaging spatial resolution of the nanocrystal arrays is determined by the distance between s_{11} and s_{12}. We

next integrated a thin film of perovskite nanocrystal arrays into a digital camera equipped with a colour CCD (Fig. 2g). The CCD has a photosensitive area of 10 mm × 14 mm and a pixel size of 2.5 $\mu m^2 \times 2.5\ \mu m^2$; the pixel size of a single azimuth detector is 50 $\mu m^2 \times 50\ \mu m^2$.

A direct application of the light-field sensor based on pixelated perovskite nanocrystal arrays is 3D imaging and light detection and ranging (Fig. 3a). This imaging system is based on a triangulation method and consists of a multiline structured light source, two lenses for light collection and a colour CCD coated with a thin film of nanocrystal arrays. The object distance z is determined by measuring the angle of the light reflected to the nanocrystal arrays by the object—that is, a high angular resolution provides a high depth resolution. For a given pixel size (50 $\mu m^2 \times 50\ \mu m^2$), the theoretical depth resolution and detectable range are improved by about 10 times and 3 times, respectively, compared with conventional triangulation methods (Supplementary Information sections 5–7). To improve data accuracy, these nanocrystal arrays were calibrated first and then the imaging system was calibrated (Fig. 3b and Supplementary Information sections 8 and 9). Under light incidences from different angles θ and ϕ, images captured using a perovskite nanocrystal array serve as a corresponding map of the colour response of each azimuth detector and the angle of incident light. To quantitatively evaluate the imaging performance of our prototype, we measured its depth accuracy as a function of scene depth and radial position within the field of view (Fig. 3c). These measurements showed an optimal depth accuracy of approximately 0.5 mm at a distance of 0.5 m, although the depth accuracy slightly decreased to approximately 1.5 mm at a distance of 2 m. Detector depth accuracy is affected by the power and angle of incident light. To ensure a high angular resolution, the power of the structured light source must be sufficient. The depth accuracy also varies depending on the intensity of the background light when the detector is used in natural-light or artificial-light conditions (Supplementary Information section 10). The dimensions of objects imaged by the light-field sensor at different distances (0.7 m and 1.5 m) are in accordance with their actual dimensions (Fig. 3d,e). Image reconstruction is also possible for objects with fine structures, such as keyboards and combs (Fig. 3f and Supplementary Information section 11). Insufficient light returned or random noise may result in undetected pixels. Moreover, we obtained 3D images of several objects of varying colours,

sizes and materials at increasing depths through pixelated colour conversion (Supplementary Information section 11).

Fig. 3: Three-dimensional imaging of real scenes by pixelated colour conversion.

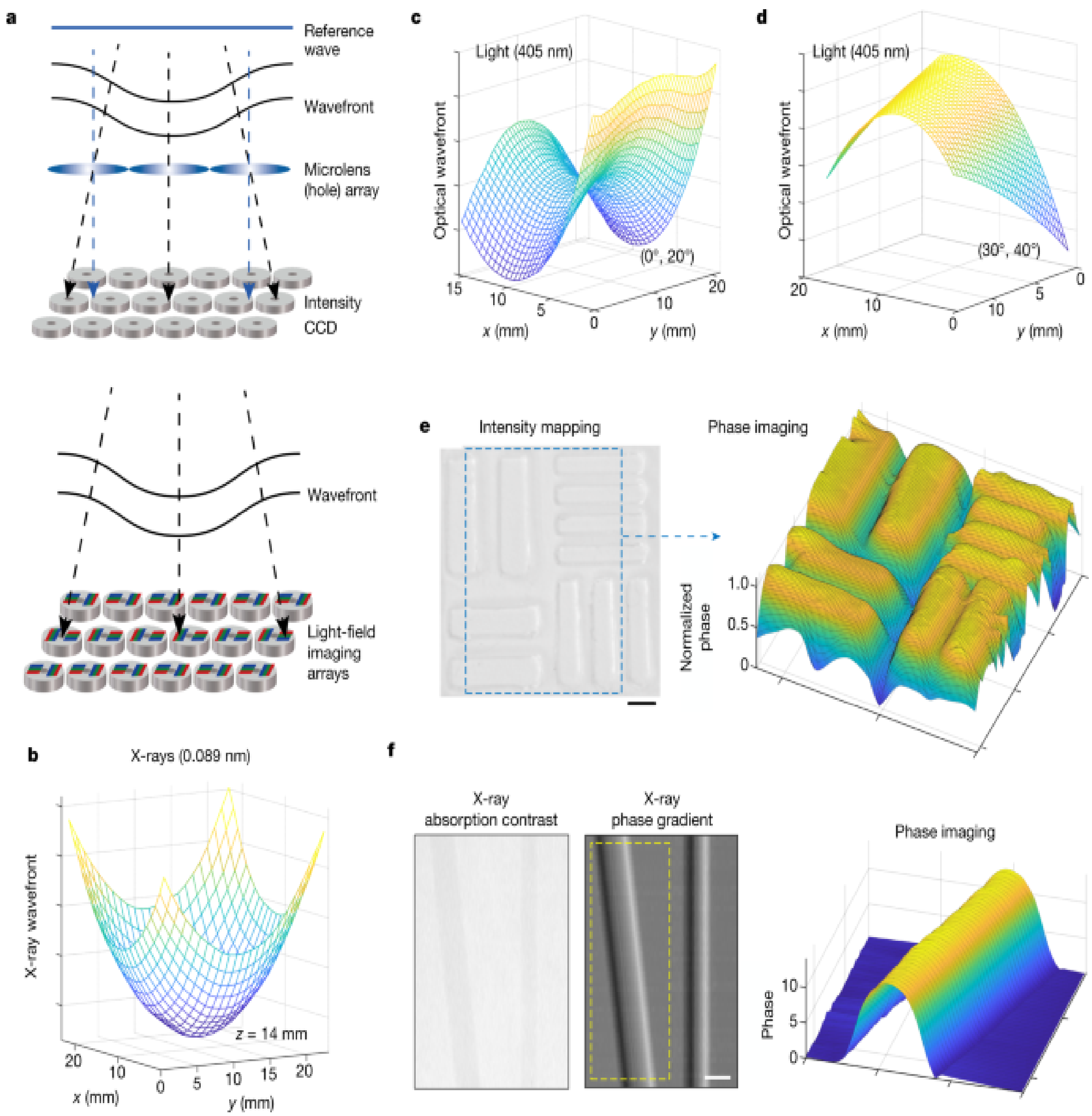


a, Schematic of the experimental set-up. Multiline structured light is incident on the object; lens 1 and lens 2 capture the reflected light and pass it to perovskite nanocrystal arrays. A colour CCD then measures the colour of

each azimuth detector to calculate the corresponding distance to the scenes. **b**, Representative images of perovskite nanocrystal arrays with incident light from different directions. **c**, Mean depth precision plotted as a function of scene depth and radial position in the field of view. A movable, flat, white screen is used as the target object. Ten measurement trials were made for each projection angle and 20 measurement trials were performed for each depth. Data are mean ± s.e.m. **d**,**e**, 3D images of scenes placed at 0.7 m and 1.5 m. **f**, 3D depth image of a keyboard captured using the 3D light-field sensor. The colour map indicates the distance from the imaging point to the z axis at the origin ($x = 0$, $y = 0$). Scale bar, 150 µm (**b**).

Another important application of pixelated colour conversion is phase-contrast imaging across a broad wavelength range from X-rays to visible light (0.002–550 nm). In phase-contrast imaging with a conventional Shack–Hartmann wavefront sensor, arrays of microlenses record the angle of incidence onto a series of grid points that determine the wavefront (Fig. 4a and Supplementary Information sections 12 and 13). A nanocrystal-array-based light-field sensor can directly measure the specific angle of visible light or X-rays to reconstruct the wavefront without microlens arrays. We first characterized the diverging wavefront of a hard X-ray beam by placing the light-field sensor 14 mm from the X-ray source (Fig. 4b). The measured wavefront curvature agrees well with analytical calculations; the maximum angle measured by the light-field sensor is 40.6°. We also mapped the visible light wavefronts in the image plane when a lens was illuminated by visible light at two different field angles (Fig. 4c,d). Furthermore, phase-contrast imaging was performed using visible light on polydimethylsiloxane (PDMS) patterns and X-rays on commercial polymethylmethacrylate (PMMA) rods (Fig. 4e,f). Surface structures can be seen in greater detail by phase-contrast imaging than by absorption-contrast imaging.

Fig. 4: Phase-contrast imaging of X-rays (0.089 nm) and visible light (405 nm) by pixelated colour conversion.



a, Principles of the Hartmann or Shack–Hartmann wavefront imaging (top) and wavefront imaging based on our 3D light-field sensor arrays (bottom). **b**, Measurement of a diverging wavefront of X-rays 14 mm from the X-ray source. **c**,**d**, Wavefronts measured in the image plane when a lens is illuminated by visible light at ($F_x = 0°$, $F_y = 20°$) and ($F_x = 30°$, $F_y = 40°$) field angles, respectively. F_x and F_y represent field angles in x and y directions, respectively. The focal length and aperture of the lens are 60 mm and 25.4 mm, respectively. **e**, Optical intensity image of a patterned PDMS substrate (left) and its phase profile (right) measured with the 3D light-field sensor. The thickness of surface structures is 0.6 mm and the wavelength of

the laser is 405 nm. **f**, Absorption contrast image of two commercial PMMA rods (1 mm and 2 mm diameters; 50 kV X-rays) and the corresponding phase-gradient map measured using the 3D light-field sensor. Phase imaging of the yellow dashed-box region is achieved using Supplementary equation (32) of the Supplementary Information. Scale bar, 1 mm (**e**).

We further compared light-field sensors based on patterned nanocrystal arrays with conventional Shack–Hartmann sensors based on microlens arrays. The fabrication of nanocrystal light-field sensors is highly robust with high uniformity over a large area compared with microlens array fabrication. In our experiment, the spatial sampling density is 400 points per mm^2, the angular resolution is approximately 0.015° and the dynamic angular range is greater than 80°. By contrast, Shack–Hartmann sensors with the same sampling density typically have a dynamic range of less than 2°. The nanocrystal light-field sensor can be used for a wider spectral range. However, microlens array sensors enable better visible-light collection because of their focusing effect. Compared with the spot localization method with Shack–Hartmann sensors, which is influenced by the quality of the spot and random noise, measuring the colour ratio in nanocrystal light-field sensors can be more robust.

In conclusion, we have presented a pixelated colour conversion strategy based on perovskite nanocrystal arrays for 3D light field detection, absolute spatial positioning, 3D imaging and visible light and X-ray phase-contrast imaging. With its current design, we have achieved a vector sensitivity of around 0.0018° and a wavelength response range of 0.002–550 nm. Although X-ray speckle-tracking methods can achieve an angular resolution close to nanoradians, coherent X-ray sources such as synchrotron radiation are a stringent requirement that cannot easily be met in many experimental and medical applications[18,19]. Further improvement in angular precision is possible by integrating high-end colour detectors. For example, a 30-bit colour display with 10-bit colour depth can yield 1.07 billion possible combinations. With advanced lithography methods and state-of-the-art processing, azimuth detector densities in excess of 10^4 pixels per mm^2 might be achieved, which could greatly improve spatial resolution of images. At present, it is not possible to detect the light field beyond 550 nm with high angular resolution using perovskite nanocrystals. However, light-field

detection and imaging through pixelated colour conversion can be readily extended to other optical materials. Sn-based perovskite nanocrystals, upconversion nanoparticles that are responsive to the near-infrared light or black phosphorus with tunable bandgaps can expand angular detection to the near-infrared, and even micrometre, wavelength range[40,41]. Furthermore, compared with Shack–Hartmann sensors, light-field sensors based on nanocrystal arrays can be directly integrated into the on-chip optical systems to measure wavefronts or phase. Because azimuth detectors can distinguish only the average vector direction of incident light, not the light from multiple directions like a light-field camera, our light-field sensors measure the average vector direction of light at each pixel. As with light-field cameras, nanocrystal light-field sensors must balance between angular and spatial resolution. Scanning light-field imaging systems can be coupled with nanocrystal arrays to further improve spatial resolution[42]. Nonetheless, the ability to image the wavefront of high-energy X-rays provides powerful solutions for optics testing and beam characterization, which can be used in applications ranging from phase-contrast imaging to gravitational-wave detection.

Methods

Chemicals

Caesium carbonate (Cs_2CO_3, 99.9%), lead(II) chloride ($PbCl_2$, 99.99%), lead(II) bromide ($PbBr_2$, 99.99%), lead(II) iodide (PbI_2, 99.99%), oleylamine (technical grade 70%), oleic acid (technical grade 90%), 1-octadecene (technical grade 90%) and cyclohexane (chromatography grade 99.9%) were purchased from Sigma-Aldrich. A Sylgard 184 silicone elastomer kit was purchased from Dow Corning for the preparation of PDMS substrates. $ZnS:Cu^{2+}/Mn^{2+}$ and $SrAl_2O_4:Eu^{2+}/Dy^{3+}$ phosphors were purchased from Xiucai Chemical.

Synthesis and characterization

$CsPbX_3$ (X = Cl, Br or I) perovskite nanocrystals were synthesized according to a previously described method[34]. First, caesium oleate was synthesized as

a caesium precursor; then $CsPbX_3$ perovskite nanocrystals were synthesized using the modified hot-injection method (Supplementary Information section 1).

Transmission electron microscopy of the synthesized perovskite nanocrystals was performed using an FEI Tecnai G20 transmission electron microscope with an accelerating voltage of 200 kV. Under visible-light or X-ray excitation, perovskite quantum dots give off narrow and colour-tunable visible emissions. Photoluminescence and radioluminescence spectra were obtained using an Edinburgh FS5 fluorescence spectrophotometer (Edinburgh Instruments) equipped with a miniature X-ray source (Amptek). An advantageous property of perovskites as detectors is their linear response to X-ray dose rate or excitation light power with coverage up to several orders of magnitude. The lowest detectable dose rate for X-ray detection is 10.8 nGy s^{-1} and the lowest detectable power for optical detection is 1 pW mm^{-2}. Perovskite quantum dots also exhibit a very fast response (decay time, $\tau = 10.4$ ns) to pulsed excitation. These nanocrystals show high photostability under successive or repeated cycles of X-ray irradiation and photoexcitation.

Principle of the 3D light-field sensor

In spherical coordinates, a detector placed parallel to the *y* axis can measure the angle variation of light around the *y* axis in the XOZ plane (Supplementary Fig. 8). For a detector placed parallel to the *x* axis, it can measure the angle variation of light around the *x* axis in the YOZ plane. Accordingly, for a beam incident from any direction (θ, φ), detector 1 placed parallel to the *x* axis detects the angle α_1 between the projection of the beam onto the YOZ plane and the *z* axis, whereas detector 2 placed parallel to the *y* axis detects the angle α_2 between the projection of the beam onto the XOZ plane and the *z* axis (Supplementary Fig. 9). The relationships between α_1, α_2 and θ, φ are as follows:

$${\alpha }_{1}=\arctan \left(\frac{y}{z}\right)=\arctan \left(\frac{r\,\sin \,\theta \,\sin \,\phi }{r\,\cos \,\theta }\right)=\,\arctan (\tan \,\theta \,\sin \,\phi)$$
(1)

$${\alpha }_{2}=\arctan \left(\frac{x}{z}\right)=\arctan \left(\frac{r\,\sin \,\theta \,\cos \,\phi }{r\,\cos \,\theta }\right)=\,\arctan (\tan \,\theta \,\cos \,\phi)$$
(2)

where α_1 and α_2 are encoded for the colour output of detectors 1 and 2, respectively. In a specific experiment, α_1 and α_2 are obtained from the CIE tristimulus value of the colour output of detectors 1 and 2, respectively. The φ and θ values of the beam are then obtained from equations (1) and (2) as follows:

$$\theta =\arctan (\sqrt{{\tan }^{2}{\alpha }_{1}+{\tan }^{2}{\alpha }_{2}})$$
(3)
$$\phi =\arctan \left(\frac{\tan \,{\alpha }_{1}}{\tan \,{\alpha }_{2}}\right)$$
(4)

Fabrication and integration of 3D light-field sensor arrays

The 3D light-field sensor based on pixelated perovskite nanocrystal arrays was fabricated using a simple moulding process (Supplementary Information section 4). First, the pre-patterned Si template was sufficiently washed with heptane. A mixture of SYLGARD silicone elastomer 184 or other silicone elastomer, curing agent and quantum dots or other luminescent materials was prepared. Then, the prepared red-emitting quantum-dot–PDMS ink and blue-emitting quantum-dot–PDMS ink were injected into the corresponding rectangular holes on the Si template and cured at 70–90 °C for 30–60 min. The top of the injected quantum-dot ink must line up with the top of the Si template. Next, 0.5-mm thick PDMS was spun onto the surface of the Si template as an adhesive film. After curing at 70–90 °C for 30 min, the PDMS film patterned with red and blue pixels was moulded off the Si template. Similarly, green-emitting quantum-dot ink was injected into the Si template and cured at 70–90 °C for 30–60 min. Then, a layer of transparent PDMS was coated on the previously processed PDMS film printed with red and blue pixels, which was then overlaid on the green-emitting ink-injected template on the mask aligner holder. Finally, after curing at 70–90 °C for

30–60 min, a film with red, green and blue pixel arrays was obtained by a mould release process. The 3D light-field sensor was formed by integrating the processed pixelated perovskite nanocrystal array film into a colour CCD, with each angle-dependent pixel covering multiple CCD pixels. The colour CCD is a Sony ICX274AL sensor with a chip size of 10 mm × 14 mm (horizontal × vertical), providing 24-bit red, green and blue true colours.

In current 3D printed moulds, the size can be adjusted from tens of micrometres to several millimetres. Large-scale manufacturing is possible through repeated demoulding. This method eliminates the need for complex semiconductor processes and special gases, which greatly reduces processing costs. Fabrication errors typically include random defects and alignment errors (Supplementary Fig. 11 and Supplementary Information section 4). Because the demoulding process used in this work has high machining accuracy and edge defects can be controlled within 0.1%, the random defect error of the entire azimuth detector pixel is almost negligible. A layer-alignment error occurs during processing because of the need to align the top and bottom layers. In cases in which the image distance is much greater than the thickness of a single-colour pixel, alignment deviation will not affect angle measurement.

Calibration of the 3D light-field sensor

The 3D light-field sensor based on perovskite nanocrystal arrays was calibrated under a collimated light-emitting diode light source. A motorized rotation stage (Daheng Optics, GCD-011060M), a pitch platform and a linear stage were connected to rotate the 3D light-field sensor in θ and ϕ directions. The image sensor was attached to the pitch platform. The pitch platform moves in the θ direction and the rotatory stage moves in the ϕ direction. A linear stage is used to compensate the off-axis movement of the image sensor when it rotates in θ direction. We present selected raw colour images captured during the calibration process to illustrate the working principle of 3D light-field sensor (Supplementary Information section 9). Each panel shows a cropped region of the raw colour image taken at different incident angles. We can observe the angular dependence of the colour output of each pixel. The yellow square represents a vertical angle-sensitive unit and the red square represents a horizontal angle-sensitive unit.

When θ increases from −40° to 40° with $\varphi = 0°$, the blueness of the pixel in the yellow square gradually becomes weaker. When θ increases from −40° to 40° with $\varphi = 90°$, the blueness of the pixel in the red square becomes gradually weaker. Once the calibration is completed for the entire range of θ and ϕ, the captured raw images are used as a lookup table for individual angle detection pixels that determine the incident angle of light.

Three-dimensional imaging procedure

The custom-built optical set-up for 3D imaging consists of a light source and an optical grating to generate multiline structured light on the 3D scene to be imaged (Supplementary Information section 11). The reflected and/or scattered light from the object is collected by a custom-made objective, which consists of two lenses optimized for focal lengths of 100 mm and 25 mm, respectively, enabling maximum angular variation for different object distances identified by the developed high-resolution 3D light-field sensor. In a typical experiment, the parameters of the multiline structured light and the camera as well as the relative distance between them were first calibrated. Next, the colour output of each detector was mapped to the angular arrays according to the calibrated results; then we can calculate the spatial coordinates x, y and z of the object point corresponding to each angular detection unit (Supplementary equations (13)–(17) in the Supplementary Information).

Spherical X-ray wavefront measurement

A 3D light-field sensor was used to measure the wavefront of spherical hard X-rays (14 keV) (Supplementary Information section 13). The X-ray source produces a divergent beam with a divergence angle of approximately 90° with wavefronts measured at different positions. The farther the 3D light-field sensor is from the X-ray source, the smaller is the measured radius of curvature of the spherical wavefront. Furthermore, from the reconstructed wavefront and slope mapping we can identify the tilt angle between the X-ray source and the sensor (Supplementary Figs. 28 and 29). As a proof of principle, colour data mapping at $z = 5$ mm was used as calibration data for reconstructing wavefronts at other distances z, resulting in less-accurate

slope measurements with increasing z. In practical applications, sampling angles must be obtained in sufficient numbers to ensure angular resolution.

Phase-contrast imaging procedure

In phase-contrast imaging using collimated UV–visible light, the object is a patterned PDMS substrate with a strip thickness of 0.6 mm. The light-field imaging sensor is placed directly behind the object to capture an image of the changed wavefront. To obtain a nearly collimated beam for X-ray phase-contrast imaging, a copper-column collimator is positioned behind the radiation source. Two commercial PMMA rods of 1 mm and 2 mm diameter are placed behind the radiation source and the light-field imaging sensor detects the changed wavefront. Specifically, the light-field imaging sensor acquires pixelated beam angles, which characterize the phase-gradient distribution. After performing median filtering and integration on the phase gradient, phase mapping can be achieved.

Data availability

The data that support the findings of this study are available in the Article and its Supplementary Information, and are also available at GitHub (https://github.com/yly1994/color-coded-light-field-imaging.git). Any further data can be obtained from the corresponding author upon reasonable request.

Code availability

Relevant codes are available at GitHub (https://github.com/yly1994/color-coded-light-field-imaging.git).

References

1. Stellinga, D. et al. Time-of-flight 3D imaging through multimode optical fibers. *Science* **374**, 1395–1399 (2021).

2. Lin, R. J. et al. Achromatic metalens array for full-colour light-field imaging. *Nat. Nanotechnol.* **14**, 227–231 (2019).

3. Robinson, I. & Harder, R. Coherent X-ray diffraction imaging of strain at the nanoscale. *Nat. Mater.* **8**, 291–298 (2009).

4. Goossens, S. et al. Broadband image sensor array based on graphene–CMOS integration. *Nat. Photon.* **11**, 366–371 (2017).

5. Gao, Y. et al. Photon-trapping microstructures enable high-speed high-efficiency silicon photodiodes. *Nat. Photon.* **11**, 301–308 (2017).

6. Gu, L. et al. A biomimetic eye with a hemispherical perovskite nanowire array retina. *Nature* **581**, 278–282 (2020).

7. Ou, X. et al. High-resolution X-ray luminescence extension imaging. *Nature* **590**, 410–415 (2021).

8. Deumel, S. et al. High-sensitivity high-resolution X-ray imaging with soft-sintered metal halide perovskites. *Nat. Electron.* **4**, 681–688 (2021).

9. Bao, J. & Bawendi, M. G. A colloidal quantum dot spectrometer. *Nature* **523**, 67–70 (2015).

10. Chen, W. et al. Continuous-wave frequency upconversion with a molecular optomechanical nanocavity. *Science* **374**, 1264–1267 (2021).

11. Jiang, Y., Karpf, S. & Jalali, B. Time-stretch LiDAR as a spectrally scanned time-of-flight ranging camera. *Nat. Photon.* **14**, 14–18 (2019).

12. Lien, M.-B. et al. Ranging and light field imaging with transparent photodetectors. *Nat. Photon.* **14**, 143–148 (2020).

13. Park, J. et al. All-solid-state spatial light modulator with independent phase and amplitude control for three-dimensional LiDAR applications. *Nat. Nanotechnol.* **16**, 69–76 (2021).

14. Kim, I. et al. Nanophotonics for light detection and ranging technology. *Nat. Nanotechnol.* **16**, 508–524 (2021).

15. Barty, A. et al. Ultrafast single-shot diffraction imaging of nanoscale dynamics. *Nat. Photon.* **2**, 415–419 (2008).

16. Shapiro, D. A. et al. Chemical composition mapping with nanometre resolution by soft X-ray microscopy. *Nat. Photon.* **8**, 765–769 (2014).

17. Pfeiffer, F., Weitkamp, T., Bunk, O. & David, C. Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nat. Phys.* **2**, 258–261 (2006).

18. Morgan, A. J. et al. Ptychographic X-ray speckle tracking with multi-layer Laue lens systems. *J. Appl. Crystallogr.* **53**, 927–936 (2020).

19. Berujon, S., Ziegler, E. & Cloetens, P. X-ray pulse wavefront metrology using speckle tracking. *J. Synchrotron. Radiat.* **22**, 886–894 (2015).

20. Fan, J. et al. Video-rate imaging of biological dynamics at centimetre scale and micrometre resolution. *Nat. Photon.* **13**, 809–816 (2019).

21. Rogers, C. et al. A universal 3D imaging sensor on a silicon photonics platform. *Nature* **590**, 256–261 (2021).

22. Zhang, X., Kwon, K., Henriksson, J., Luo, J. & Wu, M. C. A large-scale microelectromechanical-systems-based silicon photonics LiDAR. *Nature* **603**, 253–258 (2022).

23. Shen, Y. et al. Optical broadband angular selectivity. *Science* **343**, 1499–1501 (2014).

24. Yi, S. et al. Subwavelength angle-sensing photodetectors inspired by directional hearing in small animals. *Nat. Nanotechnol.* **13**, 1143–1147 (2018).

25. Fan, P., Yu, Z., Fan, S. & Brongersma, M. L. Optical Fano resonance of an individual semiconductor nanostructure. *Nat. Mater.* **13**, 471–475 (2014).

26. Cao, L. et al. Engineering light absorption in semiconductor nanowire devices. *Nat. Mater.* **8**, 643–647 (2009).

27. Kogos, L. C. et al. Plasmonic ommatidia for lensless compound-eye vision. *Nat. Commun.* **11**, 1637 (2020).

28. Pan, Q. et al. Omnidirectional photodetectors based on spatial resonance asymmetric facade via a 3D self-standing strategy. *Adv. Mater.* **32**, 1907280 (2020).

29. Jahani, S. & Jacob, Z. All-dielectric metamaterials. *Nat. Nanotechnol.* **11**, 23–36 (2016).

30. Polo, A., Kutchoukov, V., Bociort, F., Pereira, S. F. & Urbach, H. P. Determination of wavefront structure for a Hartmann wavefront sensor using a phase-retrieval method. *Opt. Express* **20**, 7822–7832 (2012).

31. Hu, L., Hu, S., Gong, W. & Si, K. Learning-based Shack-Hartmann wavefront sensor for high-order aberration detection. *Opt. Express* **27**, 33504–33517 (2019).

32. Sun, K. et al. Three-dimensional direct lithography of stable perovskite nanocrystals in glass. *Science* **375**, 307–310 (2022).

33. Kovalenko, M. V., Protesescu, L. & Bodnarchuk, M. I. Properties and potential optoelectronic applications of lead halide perovskite nanocrystals. *Science* **358**, 745–750 (2017).

34. Liu, M. et al. Colloidal quantum dot electronics. *Nat. Electron.* **4**, 548–558 (2021).

35. Sutherland, B. R. & Sargent, E. H. Perovskite photonic sources. *Nat. Photon.* **10**, 295–302 (2016).

36. Chen, Q. et al. All-inorganic perovskite nanocrystal scintillators. *Nature* **561**, 88–93 (2018).

37. Yang, Z. et al. Single-nanowire spectrometers. *Science* **365**, 1017–1020 (2019).

38. Gong, Z. et al. Full-spectrum persistent luminescence tuning using all-inorganic perovskite quantum dots. *Angew. Chem. Int. Ed.* **58**, 6943–6947 (2019).

39. Nam, T. W. et al. Thermodynamic-driven polychromatic quantum dot patterning for light-emitting diodes beyond eye-limiting resolution. *Nat. Commun.* **11**, 3040 (2020).

40. Xie, L. et al. Broadband detection of X-ray, ultraviolet, and near-infrared photons using solution-processed perovskite-lanthanide nanotransducers. *Adv. Mater.* **33**, 2101852 (2021).

41. Kim, H. et al. Actively variable-spectrum optoelectronics with black phosphorus. *Nature* **596**, 232–237 (2021).

42. Wu, J. et al. An integrated imaging sensor for aberration-corrected 3D photography. *Nature* **612**, 62–71 (2022).

Acknowledgements

This work was supported by the NUS NANONASH Program (NUHSRO/2020/002/413 NanoNash/LOA; R143000B43114) and the National Research Foundation, Prime Minister's Office, Singapore, under its Competitive Research Program (award no. NRF-CRP23-2019-0002) and under its NRF Investigatorship Programme (award no. NRF-NRFI05-2019-0003) and the RIE2025 Manufacturing, Trade and Connectivity (MTC) Programmatic Fund (award no. M21J9b0085).

Author information

Author notes

1. These authors contributed equally: Luying Yi, Bo Hou

Authors and Affiliations

1. Department of Chemistry, National University of Singapore, Singapore, Singapore

   Luying Yi, Bo Hou, He Zhao & Xiaogang Liu

2. Joint School of National University of Singapore and Tianjin University, Fuzhou, China

   He Zhao & Xiaogang Liu

3. Center for Functional Materials, National University of Singapore Suzhou Research Institute, Suzhou, China

   Xiaogang Liu

4. Institute of Materials Research and Engineering, Agency for Science, Technology and Research, Singapore, Singapore

   Xiaogang Liu

Contributions

L.Y., B.H. and X.L. conceived and designed the project. X.L. supervised the project. L.Y. characterized the materials and conducted numerical simulations. B.H. and H.Z. performed microdevice fabrication. B.H. and L.Y. performed luminescence measurements and experimental validation. B.H. contributed to the device design. B.H. and L.Y. wrote the paper and X.L. edited it. All authors participated in the discussion and analysis of the paper.

Corresponding author

Correspondence to Xiaogang Liu.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Qionghai Dai, Wanyi Nie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary Information

This file contains a table of contents leading and sections 1–14. It includes additional information on material properties, the principle of light-direction detection, device processing, design theory and the experimental apparatus for 3D imaging, system calibration methods, additional imaging data and application demonstrations to support the conclusions of the paper.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or

exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Yi, L., Hou, B., Zhao, H. *et al.* X-ray-to-visible light-field detection through pixelated colour conversion. *Nature* **618**, 281–286 (2023). https://doi.org/10.1038/s41586-023-05978-w

- Received: 14 March 2022
- Accepted: 20 March 2023
- Published: 10 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05978-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

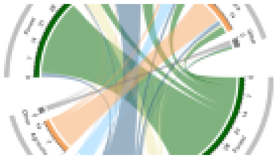
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Detectors that encode angles of incoming light as colour

Nature Research Briefing 10 May 2023

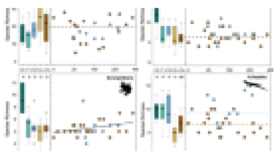
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05978-w



Trade-offs between multifunctionality and profit in tropical smallholder landscapes

04 March 2020

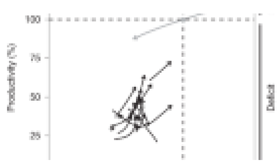
Ingo Grass, Christoph Kubitza, … Meike Wollni



Win-win opportunities combining high yields with high multi-taxa biodiversity in tropical agroforestry

26 July 2022

Annemarie Wurz, Teja Tscharntke, … Ingo Grass



Deficits of biodiversity and productivity linger a century after agricultural abandonment

28 October 2019

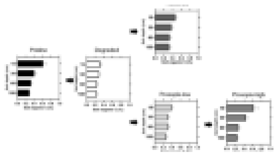
Forest Isbell, David Tilman, … Adam Thomas Clark



[Enhancing the ecological value of oil palm agriculture through set-asides]

06 February 2023

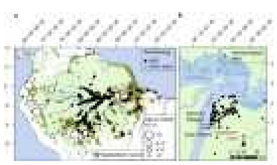
Jake E. Bicknell, Jesse R. O’Hanley, … Zoe G. Davies



[Restoration of degraded grasslands, but not invasion by Prosopis juliflora, avoids trade-offs between climate change mitigation and other ecosystem services]

24 November 2020

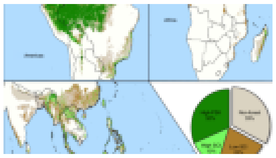
Purity Rima Mbaabu, Daniel Olago, … Urs Schaffner



[The legacy of 4,500 years of polyculture agroforestry in the eastern Amazon]

23 July 2018

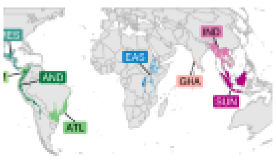
S. Yoshi Maezumi, Daiana Alves, … José Iriarte



[A policy-driven framework for conserving the best of Earth's remaining moist tropical forests]

10 August 2020

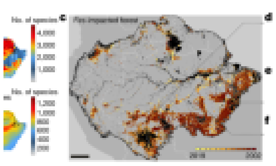
Andrew J. Hansen, Patrick Burns, … Dolors Armenteras



[Effectiveness of protected areas in conserving tropical forest birds]

14 September 2020

Victor Cazalis, Karine Princé, … Ana S. L. Rodrigues



[How deregulation, drought and increasing fire impact Amazonian biodiversity]

01 September 2021

Xiao Feng, Cory Merow, … Brian J. Enquist

- Article
- Open Access
- Published: 24 May 2023

Tree islands enhance biodiversity and functioning in oil palm landscapes

- Delphine Clara Zemp ,
- Nathaly Guerrero-Ramirez,
- Fabian Brambach,
- Kevin Darras,
- Ingo Grass,
- Anton Potapov,
- Alexander Röll,
- Isabelle Arimond,
- Johannes Ballauff,
- Hermann Behling,
- Dirk Berkelmann,
- Siria Biagioni,
- Damayanti Buchori,
- Dylan Craven,
- Rolf Daniel,
- Oliver Gailing,
- Florian Ellsäßer,
- Riko Fardiansah,
- Nina Hennings,
- Bambang Irawan,
- Watit Khokthong,
- Valentyna Krashevska,
- Alena Krause,
- Johanna Kückes,
- Kevin Li,
- Hendrik Lorenz,
- Mark Maraun,
- Miryam Sarah Merk,
- Carina C. M. Moura,
- Yeni A. Mulyani,

- Gustavo B. Paterno,
- Herni Dwinta Pebrianti,
- Andrea Polle,
- Di Ajeng Prameswari,
- Lena Sachsenmaier,
- Stefan Scheu,
- Dominik Schneider,
- Fitta Setiajiati,
- Christina Ani Setyaningsih,
- Leti Sundawati,
- Teja Tscharntke,
- Meike Wollni,
- Dirk Hölscher &
- …
- Holger Kreft

Show authors

Nature **volume 618**, pages 316–321 (2023)

- 17k Accesses
- 1 Citations
- 240 Altmetric
- Metrics details

Abstract

In the United Nations Decade on Ecosystem Restoration[1], large knowledge gaps persist on how to increase biodiversity and ecosystem functioning in cash crop-dominated tropical landscapes[2]. Here, we present findings from a large-scale, 5-year ecosystem restoration experiment in an oil palm landscape enriched with 52 tree islands, encompassing assessments of ten indicators of biodiversity and 19 indicators of ecosystem functioning. Overall, indicators of biodiversity and ecosystem functioning, as well as

multidiversity and ecosystem multifunctionality, were higher in tree islands compared to conventionally managed oil palm. Larger tree islands led to larger gains in multidiversity through changes in vegetation structure. Furthermore, tree enrichment did not decrease landscape-scale oil palm yield. Our results demonstrate that enriching oil palm-dominated landscapes with tree islands is a promising ecological restoration strategy, yet should not replace the protection of remaining forests.

Main

The loss of megadiverse tropical lowland rainforests has accelerated in the past decades[3], with deforestation and land-use change being largely driven by the rapid expansion of high-yielding cash crops such as oil palm[4]. Globally, oil palm plantations occupy 21 million hectares, mostly in Indonesia and Malaysia[5]. Although the expansion of oil palm has promoted economic development and improved livelihoods of smallholder farmers, it has also led to dramatic negative ecological impacts[6]. Compared with tropical lowland rainforests, species diversity in oil palm-dominated landscapes is greatly reduced[7], especially for forest-dependent species and species of conservation concern[4]. In addition, the transformation of forests to oil palm-dominated landscapes alters the functioning of ecological communities and environmental conditions, leading to a reduction of several ecosystem functions and services[7,8].

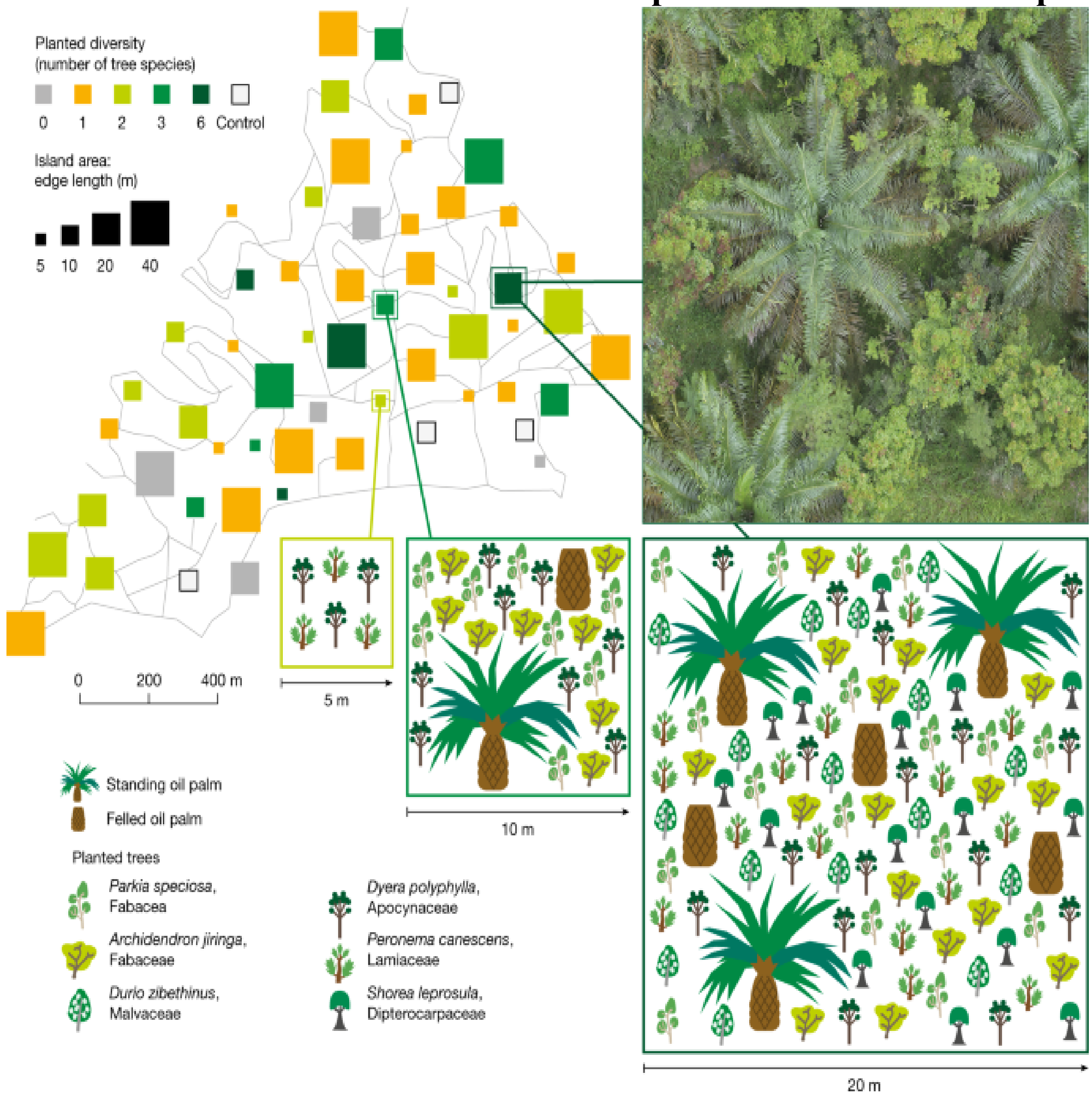
Many agricultural landscapes are in urgent need of ecological restoration to safeguard biodiversity and ecosystem functioning while also promoting local livelihoods[9,10,11], a central goal of the current United Nations decade on Ecosystem Restoration. However, trade-offs between biodiversity or ecosystem functioning and agricultural productivity may result in failed restoration efforts or lead to undesirable ecological spillover effects by promoting the expansion of the agricultural frontier into natural forested areas[12]. One way to mitigate trade-offs between restoration outcomes is to enrich agricultural landscapes with species-rich agroforestry systems[13,14] and islands of native trees through planting or natural regeneration[15,16,17]. However, to be a viable alternative for landowners, it is essential to generate empirical evidence on whether and how these restoration strategies affect

biodiversity, ecosystem functioning and agricultural productivity in cash crop-dominated landscapes[2].

Here, we present the results of a large-scale, interdisciplinary ecosystem restoration experiment, in which the restoration outcomes across 52 tree islands established in a landscape dominated by an industrial oil palm plantation (140 ha) were observed and quantified three to five years after establishment. We assessed above- and below-ground biodiversity across ten indicators representing a broad range of Kingdoms (bacteria, fungi, plants and animals; Supplementary Table 1) and 19 indicators of ecosystem functioning associated with productivity of oil palms and planted trees, resistance to invasion, pollination, soil quality, predation and herbivory, carbon and nutrient cycling and water and climate regulation (Supplementary Table 2). To provide an holistic overview of biodiversity and ecosystem functioning across the experiment, we calculated multidiversity and multifunctionality using the aforementioned indicators[18]. The experimental design allowed us to test the effects of tree island area (25, 100, 400 and 1,600 m^2) and of planted native tree diversity (zero, one, two, three and six species, with zero representing natural regeneration only) on restoration outcomes and to compare them with conventionally managed oil palm monocultures[19] (Fig. 1). Overall, we expected tree islands to enhance biodiversity and ecosystem functioning compared to conventionally managed oil palm monocultures. To provide a mechanistic understanding of the effects of planted tree diversity and island area on biodiversity and ecosystem functioning, we also measured 12 indicators of vegetation structure (Supplementary Table 3). On the basis of the theory of island biogeography[20], we expected larger tree islands to have enhanced biodiversity and ecosystem functioning compared to smaller ones. Larger tree islands potentially provide more habitats and sustain larger populations, whereas smaller islands are expected to be more like the surrounding environment, that is, the oil palm-dominated landscape. We further expected greater planted tree species diversity to favour diversity at higher trophic levels[21] and enhance ecosystem functioning through complementarity among species[22]. Planted diversity effects on restoration outcomes are probably mediated by higher vegetation structural complexity, that is, the three-dimensional distribution of plants within an ecosystem[23]. Finally, we proposed that agricultural productivity (oil palm yield) decreases at the local

scale (within tree islands), whereas the loss is negligible at the scale of the industrial plantation or landscape[16].

Fig. 1: Experimental design that tests the ecological restoration outcomes of tree island establishment in oil palm-dominated landscapes.

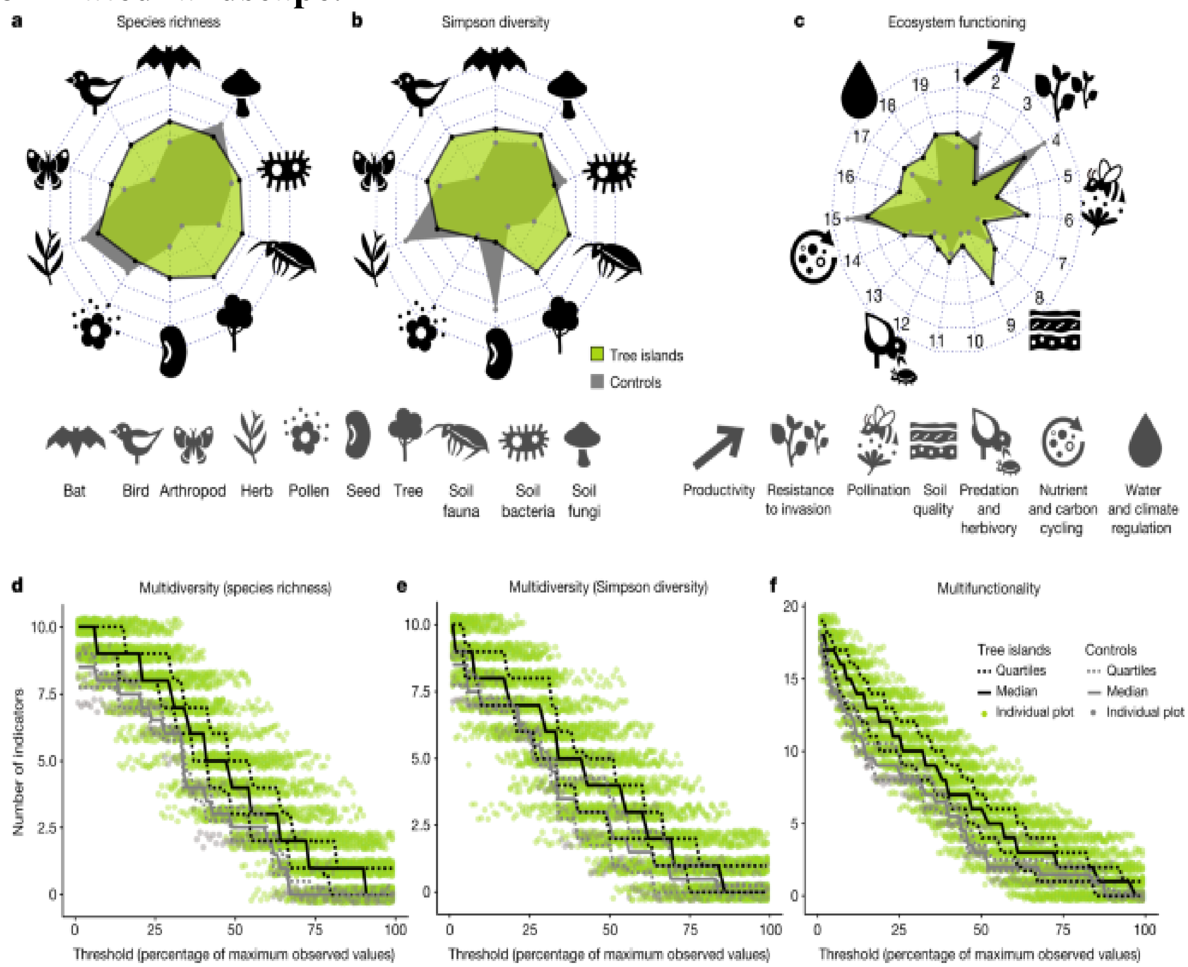


Tree islands vary in area (25–1,600 m^2) and planted tree diversity (none to six species), with a total of 52 tree islands established in an industrial oil palm plantation in Sumatra, Indonesia. Control plots represent conventionally managed oil palm monocultures. Note that the islands in the map are not at scale.

Tree islands had higher biodiversity and ecosystem functioning compared to conventionally managed oil palm monocultures (Fig. 2 and Extended Data Table 1). Yet, tree island effects on biodiversity varied depending on the indicator (tree island × indicator: $F = 2.5$, $P = 0.007$ for species richness; $F = 3.6$, P 0.0002 for Shannon diversity; and $F = 3.0$, $P = 0.001$ for Simpson diversity; Extended Data Table 1 and Extended Data Fig. 1). For example, natural regeneration and colonization led to increases in tree and bird species richness (+4.7 tree species in islands compared to monocultures,+2.5 bird species) and decreases in the diversity of the most abundant seed species (−1.2 seed species based on Simpson diversity; Supplementary Table 4). Overall, restoration benefits of tree islands were found for ecosystem functioning (tree island: $F = 6.2$, $P = 0.016$; Extended Data Table 1); with strongest increases for water infiltration (+174% saturated soil hydraulic conductivity), litter input (+151% leaf litter biomass input), activity of insectivorous bats and birds (+556%) and soil fertility (+14% 1/soil C:N ratio; Supplementary Table 5). Overall, multidiversity and ecosystem multifunctionality were higher in tree islands than in conventionally managed oil palm monocultures, regardless of the threshold used for calculation or when considering relative species abundances (Fig. 2d–f). The calculation of multidiversity (or multifunctionality) relies on the number of biodiversity (or functioning) indicators that cross a certain threshold, with thresholds expressed as the percentage of the maximum observed values within the study system[18] (here, within our landscape combining islands and conventionally managed oil palm monocultures). For example, at the 50% threshold, multidiversity increases by 1.5 in islands compared to conventionally managed oil palm monocultures. In other words, four and 2.5 biodiversity indicators reached at least 50% of their maximum observed species richness in tree islands and conventionally managed oil palm monocultures, respectively (Supplementary Table 4). Similarly, six and three ecosystem functioning indicators reached at least 50% of their maximum observed values in tree islands and conventionally managed oil palm monocultures, respectively (Supplementary Table 5). Overall, our results provide evidence of multidimensional ecological restoration benefits with tree islands in oil palm-dominated landscapes. Although the main priority is the protection of the remaining tropical forests[24], ecological restoration with tree islands along with other practices[7,25] and riparian buffer

management[26,27] plays an essential and complementary role in safeguarding biodiversity and ecosystem functioning in cash crop-dominated landscapes.

Fig. 2: Multidimensional ecological restoration outcomes in an oil palm-dominated landscape.



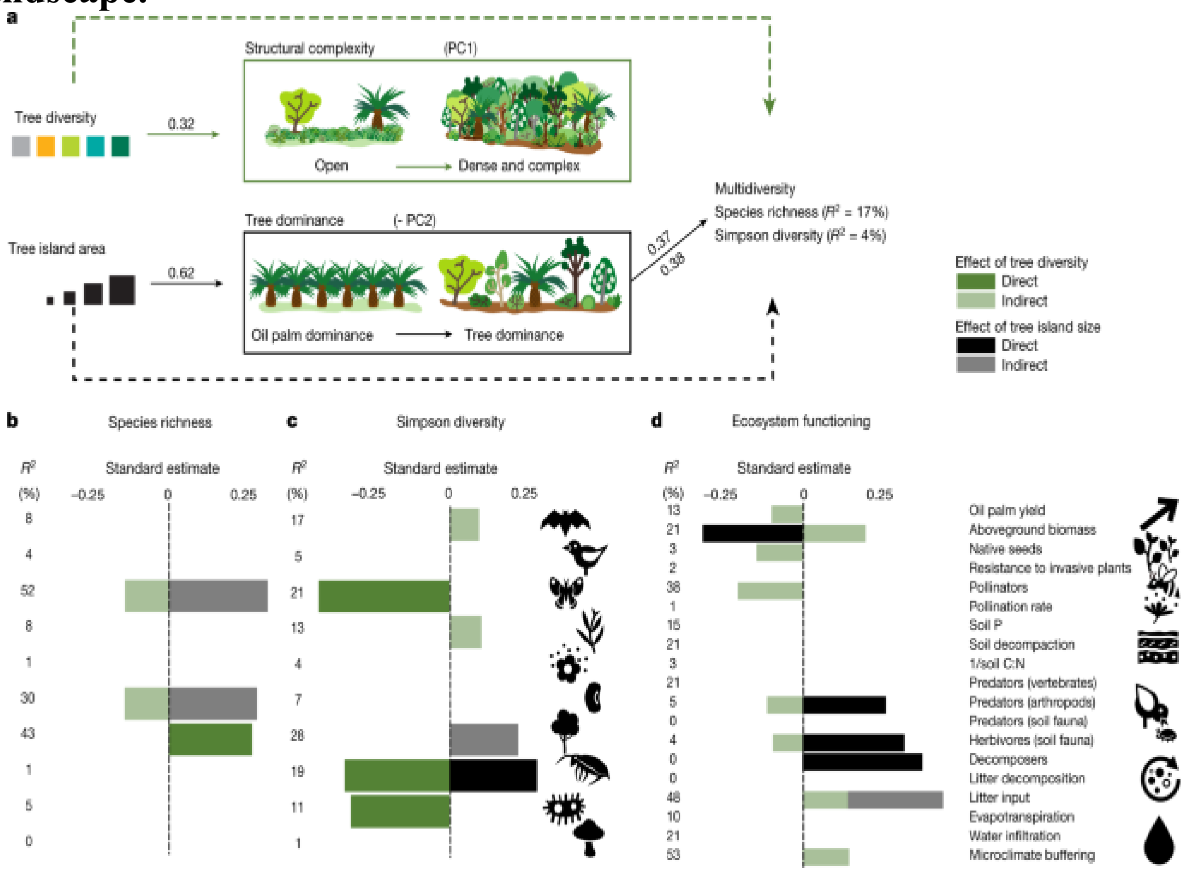
We measured 10 and 19 indicators of biodiversity and ecosystem functioning, respectively, in tree islands and compared their responses to those in plots representing conventionally managed oil palm monocultures. For ecosystem functioning, we measured: productivity as (1) oil palm yield and (2) above-ground biomass; resistance to invasion of (3) native seeds and (4) resistance to invasive plants; pollination as (5) pollinators and (6) pollination rate; soil quality as (7) soil P, (8) soil decompaction and (9) 1/soil C:N; predation and herbivory as (10) predators (vertebrates), (11) predators (arthropods), (12) predators (soil fauna) and (13) herbivores (soil fauna); carbon and nutrient cycling as (14) decomposers, (15) litter decomposition and (16) litter input; water and climate regulation as (17) evapotranspiration,

(18) water infiltration and (19) microclimate buffering. Oil palm yield (calculated per island) is considered as an ecosystem functioning because of its contribution to primary productivity, as well as agricultural productivity. **a**–**c**, Indicators of biodiversity calculated as species richness (**a**) and Simpson diversity (**b**), which emphasizes the contribution of abundant species and ecosystem functioning (**c**) across 52 tree islands (green polygons) compared to four control plots of conventionally managed oil palm monocultures (grey polygons). **d**–**f**, Polygon vertices represent median values for each indicator. Multidiversity and multifunctionality represent the number of indicators (species richness (**d**); Simpson diversity (**e**) and ecosystem functioning (**f**)) that exceed a specified threshold, which is expressed as a percentage of the maximum observed values in the oil-dominated landscape (calculated on the basis of both island and control plots combined).

Confirming our initial hypothesis, larger tree islands resulted in greater restoration benefits, both for ecosystem functioning (island area: $F = 12.9$, $P < 0.0001$; Extended Data Table 1) and biodiversity. Yet, the effects of island area on biodiversity varied across indicators (island area × indicator: $F = 5.1$, $P < 0.0001$ for species richness; $F = 2.8$, $P = 0.003$ for Shannon diversity; and $F = 1.8$, $P = 0.06$ for Simpson diversity; Extended Data Fig. 1). Our structural equation models (SEMs) revealed that the influence of the island area acted through changes in tree dominance (Fig. 3a, Extended Data Table 2 and Supplementary Tables 6 and 7), with higher multidiversity in larger tree islands that are dominated by trees rather than oil palms. The higher tree dominance in the canopy and the thicker leaf litter (Extended Data Fig. 2 and Extended Data Table 3) might provide habitats of sufficient quality and quantity to enhance multidiversity. Large tree islands may thus act as keystone structures[28] in oil palm-dominated landscapes that facilitate the arrival of seed rain (especially of locally rarer species, see Fig. 3b) and the colonization, establishment and maintenance of diverse communities, such as understorey arthropods and trees (Fig. 3b,c). Although multifunctionality also increased with island area and the effect was mediated by changes in tree dominance, the strength of the effect depended on the method used to calculate multifunctionality (Extended Data Fig. 3, Extended Data Table 2 and Supplementary Table 7). When calculated for individual functions, large tree islands were pivotal for providing ecosystem functions related to

predation and herbivory (through predatory arthropods and soil herbivores) and carbon and nutrient cycling (through decomposers; Fig 3d and Extended Data Table 3). By using constant sampling area or rarefaction curves, we could rule out that the influence of island area on biodiversity was limited to passive sampling[29] (Methods). Thus, ecological mechanisms associated with environmental filtering such as reduced edge effects and greater environmental heterogeneity probably explain the positive effects of larger tree islands on multidimensional restoration benefits.

Fig. 3: Influence of tree island area and planted tree diversity on multidimensional restoration outcomes in an oil palm-dominated landscape.



a, Effects of planted tree diversity (directly or through structural complexity) and tree island area (directly or through changes in tree dominance) on multidiversity and multifunctionality tested with SEMs. Filled arrows (and standard coefficient estimates) indicate statistically significant effects ($P < 0.05$; two-sided analysis of variance (ANOVA), $n = 52$ tree islands) and dashed arrows indicate non-significant effects. Percentage values indicate

explained variance of each endogenous variable. **b**–**d**, Effects on indicators of biodiversity quantified by species richness (**b**), Simpson diversity (**c**) and ecosystem functioning (**d**); direct effects of planted tree diversity are indicated by dark green bars and indirect effects through structural complexity are indicated by light green bars and the direct effects of tree island area are indicated by black bars and indirect effects through tree dominance are indicated by grey bars. The proportion of explained variance is shown to the left of **b** and **c**. All the bars indicate significant effects ($P < 0.05$; two-sided ANOVA, $n = 52$ tree islands). The legend for icons is presented in Fig. 2.

The effect of planted tree diversity on biodiversity—when considering abundances—depended on the biodiversity indicator (planted diversity × indicator: $F = 2.3$, $P = 0.014$ for Shannon diversity; $F = 2.8$, $P = 0.004$ for Simpson diversity; Extended Data Table 1). For example, planted tree diversity promoted the diversity of (non-planted) trees but decreased the diversity of arthropods (Extended Data Fig. 1). The statistically non-significant effects of multidiversity are probably due to contrasting responses of biodiversity indicators to planted tree diversity, with contrasting responses mediated by vegetation structure as shown by the SEM (Fig. 3b–d and Extended Data Table 2). Higher planted tree diversity led to structurally more complex habitats[30] (Extended Data Fig. 2) that benefited some biodiversity indicators (Shannon and Simpson diversity of bats and herbs), whereas others benefited from more open and structurally simpler habitats (for example, species richness of seeds and understorey arthropods; Fig. 3b,c and Supplementary Table 6). More open habitats also favoured native seeds, pollinators and predators (arthropods) and soil herbivores; whereas structural complexity enhanced above-ground biomass, litter input and microclimate buffering (Fig. 3d and Supplementary Table 7). Through changes in structural complexity, tree diversity had a negative impact on multifunctionality, although the strength of the effect depended on the methods of calculation (Extended Data Fig. 3, Extended Data Table 2 and Supplementary Table 7). Owing to specific responses associated with the adaptability of different organisms to contrasting habitats[31], establishing a combination of tree islands that differ in structural complexity may favour differences in local community composition leading to increases of gamma diversity and ecosystem functioning at the landscape scale[32].

Our study shows that the magnitude of the effect of ecological restoration on oil palm yield depends strongly on the spatial scale, with declines at the local scale, that is, within tree islands but no statistically significant reduction at the landscape scale. At the local scale, per area yields within tree islands were on average 24% lower than in the conventionally managed oil palm monocultures (Extended Data Fig. 4) because of the removal of oil palms, which reduced palm density (Extended Data Fig. 5b). In contrast, no statistically significant difference was detected in per island yield when including the yield of palms adjacent to tree islands (Fig. 2). The yield gains per oil palm surrounding the tree islands thus compensated for yield losses per area within the islands, with these beneficial effects resulting from oil palm thinning in the tree islands (Extended Data Fig. 5a,e). These beneficial effects were already observed a few years after establishment of the experiment[33] and are consistent regardless of the time period considered (Supplementary Note 1). Over time, yield decreases within tree islands are expected because of competition with trees, particularly in tree islands with higher planted diversity (Fig. 3d and Extended Data Fig. 4). Yet, these effects will remain negligible on industrial large-scale plantations because of the relatively small area covered by the tree islands. In our experiment, tree islands covered only 2.8 ha, less than 5% of the 140 ha industrial oil palm plantation. In contrast, smallholder oil palm plantations often only comprise a few hectares[5], of which (larger) tree islands would cover a more substantial proportion. In these cases, a decrease in yield may be compensated by extra goods from the tree islands, for example, fruits, natural latex, timber and firewood[34,35]. Furthermore, smallholders could benefit from higher levels of several ecosystem services, lower susceptibility to disturbance and risk diversification[17].

From an economic perspective, oil palm represents a highly profitable cash crop[6]. Consequently, replacing oil palms with native tree species typically raises concerns about high opportunity costs of lost revenue among landholders. Our large-scale study offers unique empirical evidence on the viability of multidimensional ecological benefits without compromising yield in oil palm-dominated landscapes by planting tree islands. To enhance the establishment of tree islands in oil palm- and other cash crop-dominated landscapes, they could be incorporated as a requirement in existing sustainability certifications (for example, Roundtable for Sustainable Palm

Oil), alongside other practices including optimized management[25], ecological intensification and riparian restoration[26,36,37]. Enhancing the status of sustainability certifications should, however, not come at the expense of smallholder farmers, who are often excluded from certification programmes[38], nor at the expense of the protection of remaining intact forests for their exceptional value as refugia for biodiversity and providers of ecosystem functioning[39]. Overall, we provide robust evidence that biodiversity and ecosystem functioning in cash crop-dominated tropical landscapes can be enhanced without compromising overall agricultural productivity by planting tree islands. Although our study was conducted in a single landscape, it adds to growing experimental[37,40,41] and modelling evidence[42] of the ecological and economic benefits in oil palm agroforestry systems. Understanding how biodiversity and ecosystem functioning change in several landscapes[43,44] is urgently needed for designing and scaling-up ecological restoration of oil palm landscapes worldwide.

Methods

The biodiversity enrichment experiment

Our study was conducted in EFForTS-BEE, the biodiversity enrichment experiment of the EFForTS project (Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia))[19]. EFForTS-BEE is part of the global network of tree diversity experiments TreeDivNet[45] (https://treedivnet.ugent.be/). The study region is characterized by a humid tropical climate with a mean temperature of 26.7 ± 0.2 °C and an annual rainfall of 2,235 ± 381 mm and the dominant soil type is loamy Acrisol[46]. In December 2013, 52 experimental plots (tree islands) were established in a conventionally managed 140 ha oil palm plantation. Following a random partition design[47], we systematically varied island area (25, 100, 400 and 1,600 m^2) and planted diversity (zero, one, two, three and six tree species). The six planted tree species (*Archidendron jiringa* (Jack) I.C.Nielsen (Fabaceae), *Parkia speciosa* Hassk (Fabaceae), *Durio zibethinus* L. (Malvaceae), *Dyera polyphylla* (Miq.) Steenis (Apocynaceae), *Shorea leprosula* Miq. (Dipterocarpaceae) and *Peronema canescens* Jack (Lamiaceae)) are native to the region and widely used for

their fruits, timber or latex[35]. Around 40% of the oil palms located inside the tree islands were felled, with the number of felled oil palms differing depending on the tree island area[33]. The trees were planted between the felled and standing oil palms on a 2 m triangular grid. The tree islands were fenced and the management comprised a total stop of fertilizer, herbicide and pesticide application after planting. After May 2016, manual weeding was restricted to 1 m circles around the planted trees when these were shorter than the surrounding grass layer, allowing for natural regeneration.
In addition to the 52 tree islands, we established four control plots in the oil palm plantation that have a fixed area (100 m^2) and that were managed conventionally (that is, no oil palm was felled, no tree was planted and application of fertilizer, herbicide and pesticide was as usual), in the main text referred to as conventionally oil palm monocultures. In total, the experiment comprises 56 study plots[19]. In each study plot larger than 25 m^2, one subplot of 5 × 5 m^2 was established in a random location at a minimum distance of 1.5 m from the plot edge.

Field measurements

We conducted an interdisciplinary field campaign from October 2016 to October 2018, that is, 33–57 months after establishment of the experiment. At this early stage of the experiment, the tree islands already differed in their structural complexity[30] and the planted trees reached up to 16 m height[35]. In all the 56 study plots, several indicators related to biodiversity, ecosystem functioning and structure were measured using standardized procedures and constant sampling areas at the level of the plot or subplot (Supplementary Tables 1–3). Only trees were sampled at unequal areas (that is, all trees present in the plots were sampled) and were therefore standardized using rarefaction curves (see ‘Trees’). Oil palm yield was continuously measured since the beginning of the experiment at the level of individual palm but the data were then aggregated over space and time (see ‘Per area yield’ and ‘Per island yield’). Each variable presented in the main text had one measurement per plot, such that blinding and randomization were not applicable.
No statistical methods were used to predetermine sample size. The data were processed and analysed in R v.1.2.1335 (ref. [48]).

Birds and bats

We recorded audible and ultrasound in March 2017 using automated sound recorders (SM2Bat+ recorders, Wildlife acoustics; with an acoustic SMX-II microphone on the left channel and one full-spectrum Sonitor Parus[49] microphone on the right channel), strapped to wooden poles at a height of 1.5 m in the centre of the plot. On consecutive days, we extracted sound recordings for sampling birds and insectivorous bats. We used two stereo 15 min recordings starting 15 min before sunset and two 15 min stereo recordings starting at sunrise, sampled at 22.05 kHz for birds. We used two 40 min mono sound recordings from the right channel, extracted from consecutive nights, starting 20 min after sunset, sampled at 384 kHz for bats. Twelve sound recorders were installed simultaneously in 12 randomly chosen plots. The recordings were annotated in ecoSound-web[50] to extract the duration of each bird vocalization and bat pass and bird detection distances were estimated using reference sound transmission sequences[51]. We assigned birds to species according to Birdlife International taxonomy. Owing to the lack of standard protocols and reference collections for Southeast Asia, we could not identify bats to species and used sonotypes instead. We appended feeding guild information to each bird species[52] all detectable bats were echolocating and thus considered insectivorous bats. Only bird vocalizations detected within a 28 m radius were included, which corresponds to the diameter of the largest study plot (40×40 m^2). We used the maximum number of individuals detected simultaneously in all recordings per plot as a conservative proxy of abundance per bird species or bat sonotype.

Understorey arthropods

Arthropods were sampled in the understorey vegetation during October 2016 to January 2017. Each plot was sampled three times with six pan traps per plot exposed for 45 h. Traps were made of white plastic soup bowls covered with yellow ultraviolet spray-paint[53] and were filled with water and one drop of regular soap. They were fixed in a holding system in groups of three at the height of the surrounding plants and these systems then equally distributed in distance from edge and to each other. All arthropods were preserved in 70% ethanol. Subsequently, all individuals were identified to higher taxonomic groups and morphospecies. The taxonomic groups Hymenoptera, Lepidoptera and Araneae were categorized into functional

groups (pollinators, predators and parasitoids) using different identification keys[54,55,56,57,58,59,60] (Supplementary Table 10). Predators and parasitoids were merged into the single functional group 'predators'.

Soil fauna

During October–November 2016, in each plot, four soil samples of 16 × 16 cm^2 were taken randomly within the subplot with a spade. Samples included litter (if present) and soil down to a depth of 5 cm. Animals were extracted using a gradient heat extractor[61] and collected in dimethyleneglycol–water solution (1:1) and thereafter transferred into 70% ethanol[62]. All extracted animals were counted and sorted into 28 taxonomic groups (in most cases orders) allowing for functional group classifications[63]; Extended Data Table 4. We calculated community metabolism of all animals that were classified as detritivores, herbivores and predators in a sample by using mean group- and ecosystem-specific estimates derived from ref. [63]. The estimates are based on measurements of more than 5,000 individuals of soil animals across eight different oil palm plantations in the same region; to estimate community metabolism, individual body masses were recalculated to metabolic rates using group-specific regressions from ref. [64]. Community metabolism was calculated by summing up metabolic rates of all individuals; we used the mean per plot across four samples for each functional group (detritivores, herbivores and predators) for the analysis. We also computed taxonomic diversity as the number of taxonomic groups (in most cases orders) present in each plot for the analysis.

Fungi

In January 2017 three soil cores (10 cm depth, 4 cm diameter) were taken within each 5 × 5 m^2 subplot. Surface leaf litter was removed before soil collection. The soil was sieved through a 50 × 50 mm^2 sieve and roots were separated from soil. The fungal community was assessed using Illumina next-generation sequencing (Illumina) of the ITS2 marker region. The detailed protocol for amplification, amplicon sequencing and generation of fungal operational taxonomic units (OTU) is described in ref. [65]. OTUs were

classified taxonomically using the BLAST (blastn, v.2.7.1) algorithm[66] and the UNITE v.7.2 (UNITE_public_01.12.2017.fasta) reference database[67].

Prokaryotes

In May 2017 three cores of topsoil (10 cm depth) were taken in each subplot. Soil cores were then mixed, homogenized and freed from roots. A total of 5 ml of RNAprotect Bacteria reagent (Qiagen) was added to 5 g of soils to prevent nucleic acid degradation. DNA and RNA were extracted from 1 g of soil by using the Qiagen RNeasy PowerSoil Total RNA Kit and the RNeasy PowerSoil DNA Elution Kit (Qiagen). The V3–V4 region of the 16S ribosomal RNA gene was amplified and sequenced as described in ref. [68]. Paired-end sequences were quality filtered with fastp (v.0.20.0)[69] and merged with PEAR v.0.9.11 (ref. [70]). Remaining primer sequences were clipped with cutadapt v.2.5 (ref. [71]). Size filtering, dereplication, denoising and chimaera removal was performed with vsearch v.2.12.0 (ref. [72]). Curated sequences were then classified by mapping each sequence against the SILVA database with the BLAST[73]. Counts were normalized by using the GMPR normalization[74].

Seeds

We installed four seed traps in each of the 56 study plots for 1 yr, that is, between 1 April 2017 and 29 March 2018. The traps were built using fine-mesh cloth attached to a squared structure made in PVC pipes of size 50 cm × 50 cm fixed at 1 m from the ground. The traps were installed at random locations in each of the four quadrants of each plot, at a minimum distance of 1 m from the plot edge. The contents of the traps were collected twice a month, dried at approximately 40 °C during 3–7 days. All the seeds were carefully extracted from the samples, counted and separated by morphospecies using hand lenses (×10 magnification) and a microscope (Leica photomicroscope with ×400 magnification) for very small seeds. Molecular identification of the morphospecies was implemented using three universal plant DNA barcodes (*matK, rbcL* and ITS2)[75,76,77,78] and taxonomic assignments were made using BLASTn search against the NCBI Genbank reference sequence database[79]. Sequences obtained from the

barcode loci were deposited in NCBI Genbank under the accession numbers OM811991–OM812021, OM837673–OM837724 and OM935782–OM935815. We classified each morphotype as native or non-native species using available literature[80,81] (http://www.plantsoftheworldonline.org/). We derived the native seed density (number of native seeds per m^2) as the total number of native seeds over the entire sampling duration per plot, which was used as an indicator of ecosystem functioning (see ‘Ecosystem functioning’). Seed diversity, calculated on the basis of the Hill number frameworks and used as indicators of biodiversity (see ‘Biodiversity’), was derived from the pooled samples per plot over the entire sampling duration for all seeds (native and non-native).

Herbs

All non-woody terrestrial vascular plants (for example, angiosperm herbs and vines, ferns, but not epiphytes) in the subplot were recorded from February until March 2018. They were classified to species or morphospecies and herbaceous cover (in absolute per cent ratios from 1% to 100%) was estimated by two people. Epiphytes growing on the stems of trees or palms were excluded, whereas vine species that rooted in the ground and climbed up stems of trees or palms were included. Herbarium specimens were collected and stored in the laboratory of Jambi University. All names were checked following The Plant List 2013, v.1.1 (http://www.theplantlist.org).

Trees

All planted trees were surveyed in January to February 2018 as part of a yearly inventory[35]. Furthermore, we surveyed all free-standing woody plants (trees, shrubs and bamboos) that colonized the plots with a length of ≥130 cm from April until August 2018. For each species or morphospecies, one voucher specimen was collected, dried and pressed according to standard procedure. In the main text, we refer to the colonized woody plants as ‘trees’, unless stated otherwise. Because the number of sampled trees largely varied according to the tree island area, we standardized the diversity

estimates using rarefaction curves (R package iNEXT)[82] to 24 individuals, which represent the median number of individuals per plot.

Pollen

To collect pollen/spore rain, Behling pollen traps[83] were installed from June until October 2018. Each trap consists of a plastic tube which is placed about 30 cm above the ground and is held by a fixing pole. The tube is filled with 5 ml of liquid glycerol, synthetic cotton and, on the top, it is covered by a mosquito net to reduce disturbance from animals or litter and prevent the cotton from being removed. In tropical regions heavy rainfalls occur, thus it is necessary to prevent the pollen from pouring out of the pollen trap. In the Behling trap, glycerol is used, which has a higher density compared to water. Consequently, the incoming rainfall can flow out of the trap without taking the pollen, which is trapped in the synthetic cotton and in the glycerol[83]. The Behling traps were modified to mimic the surrounding environment and maximize recovery. In total, 168 pollen traps were installed in the plots (3× plot). Of the total 56 plots, the pollen traps were not recovered in three plots (P28, P34 and P47). One pollen trap from each 53 plot was processed and analysed. Firstly, each pollen trap was washed with distilled water through a 2 mm mesh sieve to remove large size materials. Afterwards, the pollen traps were sieved through a 150 µm mesh sieve to exclude medium-sized materials from the samples. Two *Lycopodium* tablets were added as markers to each sample to estimate palynomorph concentrations[84] and the Erdtman acetolysis[85] was applied, to remove cellulose material. Residues were mounted in glycerol jelly for pollen visualization, identification and counting. Pollen and spore analyses were carried out using light microscopy. All identified pollen and spore types were photographed using a Leica photomicroscope with ×400 magnification. For each trap, a total sum of at least 100 pollen grains were counted. Pollen and spore grains can rarely be identified to species level and the level of taxonomic identification varies for different groups of plants. Consequently, a reduction to the family level has been proposed for studies involving analysis of palynological diversity in the tropics[86].

Pollination

We assessed pollination rate on chilli pepper plants (*Capsicum annuum*) as phytometer plants, selected for potential shade tolerance[87], widespread home garden cultivation in this region[88] and the potential role pollination can play in fruit quality and yield[89]. We raised 1,500 individuals of a locally available variety of *C. annuum* from seed outside of the study plots. During the growth period outside the study plots, we applied NPK fertilizer and pesticide (imidacloprid, deltamethrin, mancozeb and abamectin) following local practices to standardize growing conditions and control pest damage before transfer to field sites. We halted fertilizer and pesticide application 1 week before placement in the plots and only watered as conditions required thereafter. In February 2018, we selected 224 healthy individuals of comparable size to transfer to the 56 study plots (four plants per plot). The four chilli plants were placed, still in their pots, at the centre of each plot for 5 weeks for a period of open pollination and monitoring, followed by 3 weeks for fruit harvesting. We removed any flowers before placement in the field, so pollinated flowers and developing fruits were assumed to result from pollination within the study plots. During the period before final harvest, each plot was revisited once per week and the number of flowers were counted per plant. The rate of successful pollination was estimated from the fruit to flower ratio, which was the total number of harvested fruits divided by the total number of flowers observed per plot.

Per palm yield

We followed the conventional harvesting procedure established by the plantation manager of PT Humusindo and measured the weight of the fresh fruit bunches directly after harvest using a portable scale. We measured the per palm yield (kg per palm) of all palms inside the 52 tree islands ($N = 214$) and one palm per control plot with conventional management ($N = 4$). To obtain a more solid estimate for the conventional plantation, we measured the per palm yield of 30 more reference palms that were evenly distributed across the conventional plantation at approximately an equal distance to each tree island and whose neighbourhood is characteristic of conventionally managed oil palm monocultures (Supplementary Note 2 and Supplementary Figs. 1–3). To examine potential changes in yield in the conventionally managed oil palm plantation surrounding the tree islands (‘spillover effects’), we measured the per palm yield of three oil palms adjacent to each

tree island, at increasing distance to the island's edge (at position number 1, 2 and 3)[33]. For direct comparison with earlier findings, we analysed the data following established methodology[33]. The tests were based on linear mixed-effect models with the annual yield of individual oil palms as the response variable and the plot identity as the random effect. Pairwise comparison was conducted with a post hoc Tukey test. Because our results indicate that only the palm directly adjacent to the tree island (in position number 1) was affected by the experimental treatment (Extended Data Fig. 5a, in agreement with ref. [33]), we do not consider the palms in position 2 and 3 in the yield calculation per island (Per island yield). The per palm yields in and adjacent to the 56 study plots have been continuously monitored since the establishment of the experiment in December 2013. The extra 30 palms were established in December 2016. For consistency with other indicators (Extended Data Tables 1–3), we reported yield data for 1 yr (November 2017 until October 2018) in the main text and for 2 yr (November 2016 until October 2018) in the Supplementary Note 1. Yield data since 2014 are shown in Extended Data Fig. 4, in which the oil palms in position 3 were used as reference palms for the corresponding time period.

Per area yield

We estimated per area yield (ΔY_{ha}, kg ha^{-1}) as the yield of a given palm (kg palm^{-1}) multiplied by a stand density-dependent expansion factor (EF) to derive estimates of per area yield (kg ha^{-1}). We then calculated the per area yield change between tree islands and reference (kg ha^{-1}; Supplementary Note 3). This approach accounts for changes in per area yield due to oil palm thinning (that is, reduced oil palm densities and changes in per palm yield in the tree islands) but does not account for potential changes in per palm yield on the surrounding plantation, for example, because of spillover effects[19]. An alternative analysis considering spillover effects was performed at the plot level (Per island yield).

Per island yield

We estimated oil palm yield changes at the tree island scale ($\Delta Y_{\mathrm{Island}}$, in kg island^{-1}); equations (1)–(4) following established methodology[33]. This

method considers the yield foregone owing to the removal of some oil palms before the experiment, as well as changes in per palm yield inside the tree islands and directly adjacent to the tree islands (at position 1, that is, spillover effects). Because the number of oil palms inside and adjacent to the tree islands and the number of removed oil palms vary depending on tree island area[33], the net oil palm yield changes are provided per plot and not per area. Even though this method was initially designed to calculate the oil palm yield changes for the tree islands, here we also apply it to the four control plots to integrate them in our synthesis analysis.

$${\varDelta Y}_{{\rm{Island}}}={Y}_{{\rm{Spillover}}}+{Y}_{{\rm{RemainChange}}}-{Y}_{{\rm{Foregone}}}$$
(1)
$${Y}_{{\rm{F}}{\rm{o}}{\rm{r}}{\rm{e}}{\rm{g}}{\rm{o}}{\rm{n}}{\rm{e}}}={N}_{{\rm{f}}{\rm{e}}{\rm{l}}{\rm{l}}{\rm{e}}{\rm{d}}}\times {Y}_{{\rm{p}}{\rm{_}}{\rm{r}}{\rm{e}}{\rm{f}}}$$
(2)
$${Y}_{{\rm{R}}{\rm{e}}{\rm{m}}{\rm{a}}{\rm{i}}{\rm{n}}{\rm{C}}{\rm{h}}{\rm{a}}{\rm{n}}{\rm{g}}{\rm{e}}}={N}_{{\rm{i}}{\rm{n}}}\times ({Y}_{{\rm{p}}{\rm{_}}{\rm{i}}{\rm{n}}}-{Y}_{{\rm{p}}{\rm{_}}{\rm{r}}{\rm{e}}{\rm{f}}})$$
(3)
$${Y}_{{\rm{Spillover}}}={N}_{{\rm{adj}}}\times \left({Y}_{{\rm{p}}_{\rm{adj}}}-{Y}_{{\rm{p}}_{\rm{ref}}}\right)$$
(4)

$\Delta Y_{\rm Island}$, per island oil palm yield change (kg island^{-1})

$Y_{\rm Spillover}$, per island yield changes due to spillover effects (kg island^{-1})

$Y_{\rm RemainChange}$, per island yield changes inside the island (kg island^{-1})

$Y_{\rm Foregone}$, per island yield foregone due to oil palm removal (kg island^{-1})

$N_{\rm i}$, number of remaining oil palms inside the island

N_{ad}, number of oil palms directly adjacent to the island (that is, adjacent position 1)

N_{felled}, number of removed oil palms in the island

Y_{p_ref}, median per palm yield of the reference palms in the conventionally managed oil palm plantation (kg palm^{-1})

Y_{p_in} median per palm yield inside the tree island (kg palm^{-1})

Y_{p_adj}, per palm yield directly adjacent to the tree island (that is, adjacent position 1 (kg palm^{-1})).

Above-ground biomass

For all the planted trees, we measured the basal diameter (at 10 cm above ground), the diameter at breast height (130 cm above ground) and the tree height in 2017 as part of a yearly inventory[35]. In January and February 2017, we also measured the height of the oil palms at meristem, that is, the point of attachment of the young leaves to the oil palm trunk[30]. We estimated above-ground biomass of the trees (equation (5)) and the oil palms (equation (6)) using the respective allometric equations of refs. [90,91]:

$${{\rm{AGB}}}_{{\rm{tree}}}=0.0673\times {(\rho \times {{\rm{DBH}}}^{2}\times H)}^{0.976}$$
(5)
$${{\rm{AGB}}}_{{\rm{palm}}}=71.797\times H-7.0872$$
(6)

AGB_{tree}, above-ground biomass of the planted trees (kg tree^{-1})

AGB_{palm}, above-ground biomass of the oil palms (kg palm^{-1})

DBH, tree diameter at breast height (cm)

H, height of tree or palm (m)

ρ, wood density (g cm^{-3}).

Wood density for *Peronema canescens* (0.61 g cm^{-3}), *Parkia speciosa* (0.54 g cm^{-3}) and *Dyera polyphylla* (0.36 g cm^{-3}) was based on EFForTS core plot data, whereas for *Archidendron* sp. (0.36 g cm^{-3}), *Shorea leprosula* (0.44 g cm^{-3}) and *Durio zibethinus* (0.516 g cm^{-3}) it was taken from the global wood density database[92].

We estimated the total above-ground biomass per plot as the sum of the above-ground biomass of the palms and the planted trees (equation (7)). The estimations of total AGB did not consider the necromass, litter, understorey vegetation and spontaneously established trees, which were considered negligible.

$${\rm{A}}{\rm{G}}{\rm{B}}=(\,\sum {{\rm{A}}{\rm{G}}{\rm{B}}}_{{\rm{t}}{\rm{r}}{\rm{e}}{\rm{e}}}+\sum {{\rm{A}}{\rm{G}}{\rm{B}}}_{{\rm{p}}{\rm{a}}{\rm{l}}{\rm{m}}}\,/{N}_{{\rm{i}}{\rm{n}}}\times {d}_{{\rm{p}}{\rm{a}}{\rm{l}}{\rm{m}}}\,)/(A\times 10,000)$$
(7)

AGB, total above-ground biomass per plot (t ha^{-1})

d_{palm}, density of oil palms (number of oil palms per ha) that takes into account the local neighbourhoods of the plots (also referred to as EF; see Supplementary Note 3)

A, area of the plot (m^2).

Tree growth

The growth of the planted trees per plot was calculated as[35]:

$${{\rm{B}}{\rm{A}}}_{{\rm{i}}{\rm{n}}{\rm{c}},2017-2018}=\sum ({{\rm{B}}{\rm{A}}}_{{\rm{t}}{\rm{r}}{\rm{e}}{\rm{e}},2018}-\sum {{\rm{B}}{\rm{A}}}_{{\rm{t}}{\rm{r}}{\rm{e}}{\rm{e}},2017})/A.$$
(8)

$BA_{inc,\ 2017–2018}$, total plot-level basal area increment between 2017 and 2018 (in $cm^2\ m^{-2}\ yr^{-1}$, equivalent to $m^2\ ha^{-1}\ yr^{-1}$)

$BA_{tree,\ year}$, tree basal area (in cm^2) derived from the basal diameter (cm) in the specific year.

Leaf litter input

We measured leaf litter fall (in $g\ m^{-2}\ yr^{-1}$) using the four seed traps installed randomly in each four quadrants of the plots from April 2017 to March 2018 (Seeds). The contents of the traps were collected twice a month, dried at 40 °C for 4–7 days and weighted. We also sorted the leaves by species and weighted the content for the six planted tree species and oil palm separately.

For each sampling date, we aggregated the values at plot level using the median per plot of the litter weight. We then excluded outliers defined as plot-level values outside the range of 3 standard deviations around the median of the entire data (less than 5% of the litter weight data, total and per species). To get annual estimates, we summed the available plot-level values over time and divided them by the number of sampling dates (that is, between 17 and 24, depending on the number of missing traps or excluded outliers). We then multiplied the obtained values by the seed trap area (0.25 m^2) to get the leaf litter fall in $g\ m^2\ yr$.

Leaf litter decomposition

We installed litterbags ($20 \times 20\ cm^2$, 4 mm mesh size) each filled with 12 g of material: 6 g of freshly cut and air-dried (approximately 25 °C) fronds of oil palm leaves[93] and 6 g of the freshly fallen air-dried leaf litter for each tree species or their combinations in experimental plot. In each plot, one litterbag was installed in November 2017 for a duration of 6 months. Decomposition (litter mass loss) was calculated as the difference between the initial litter dry mass and litter dry mass remaining after 6 months and expressed as a percentage of decomposed material.

Water infiltration capacity

To quantify soil water infiltration capacity, we measured saturated soil hydraulic conductivity (K_{fs}, cm h^{-1}) using a dual-head infiltrometer (Saturo) in March 2018 near the subplot centre in 35 (out of the 52) tree islands and in the four control plots representing conventionally managed oil palm monocultures. Owing to a broken instrument, the 17 remaining plots were measured using a custom manual double-ring infiltrometer, which tends to yield higher K_{fs} estimates than the dual-head approach because there is no correction for lateral flow. In three plots, K_{fs} was measured with both devices. We plotted these values against each other and found a close linear relationship ($R^2 = 0.98$, $P = 0.066$); even though it was only marginally significant because of the small sample size, we used it to correct the values from the 17 plots that were measured manually ($K_{fs_corr} = 1.44 + 0.55 K_{fs_double_ring}$) to allow for comparability across all 56 plots.

Evapotranspiration

We recorded land and canopy surface temperatures using a radiometric thermal camera (FLIR Tau 2 640, FLIR Systems) attached to a TeAx ThermalCapture module (TeAx Technology GmbH) mounted on a multicopter drone (MK EASY Okto V3; HiSystemsy) as described in ref. [94]. Image sets were recorded four to five times per day around noon, covering each plot once over a 9 day period encompassing varying weather conditions. Land and canopy surface temperatures were the main input for modelling latent heat flux (in W m^{-2}) and deriving evapotranspiration using the QWaterModel QGIS3 Plugin[95], which is based on the DATTUTDUT energy balance model[96]. Measured short-wave radiation and relative humidity were used as further input variables to support the prediction of latent heat flux and derive evapotranspiration.

Microclimate

We measured microclimate using temperature per humidity loggers: hydrochron (DS1923-F5) and thermochron (DS1922L-F5) iButtons, Maxim integrated. The loggers were installed in the middle of each plot at 1.5 m above ground and were protected from water and direct solar radiation using

handmade multiplate radiation shields[97]. Data were collected for 1 yr (18 November 2017 until 19 September 2018) every 3 h starting at midnight. As a proxy for microclimate buffering, we calculated the daily amplitude as the absolute difference between values at 7:00 and 15:00 (ref. [97]), aggregated using the median value over the entire measurement period.

Soil properties

We determined soil total carbon content (g mg^{-1}), total nitrogen content (mg g^{-1}) and plant available phosphorus content (mg g^{-1}) using the same three soil samples as for fungi community data (see 'Fungi') and the method of determination is described in detail in ref. [65]. We then calculated the C:N ratio accounting for the molar mass of the elements following ref. [98], that is, 12.0107 for carbon and 14.0067 for nitrogen. We also measured soil bulk density (g cm^{-3}) using five soil samples taken in the subplot in May 2018. Soil rings of 100 cm^3 were inserted horizontally into the first 5 cm of topsoil. The soil was weighed, dried at 105 °C until constant weight and weighted again. Calculation was done on the dry weight basis, for which the sample dry weight (g) was divided by the volume of the sample (cm^3) collected from the average of the five replicates. We used the mean per plot for all mentioned soil variables and used the inverse of C:N ratio and soil bulk density as measures of soil fertility and soil decompaction, respectively (Supplementary Table 1).

Vegetation structure

We measured 12 variables representing various aspects of the vegetation structure (Supplementary Table 3). We used a terrestrial laser scanner Focus M70 (Faro Technologies) to create three-dimensional point clouds of the vegetation at the centre of each plot in September and October 2016, as described in ref. [30]. We computed the (1) stand structural complexity index (SSCI) following ref. [99] and its two components: (2) the mean fractal dimension index (MeanFRAC) derived from cross-sections of polygons in the three-dimensional point cloud, which is a scale-independent and density-dependent measure of structural complexity and (3) the effective number of layers (ENL) that describes vertical stratification based on the Simpson

Index[100]. ENL and MeanFRAC are integrated in the SSCI and all these three measures were derived from vegetation parts above 130 cm. We also derived (4) the understory complexity index that measures the fractal dimension of horizontal cross-sections of the point cloud between 80 and 180 cm height, thereby measuring the structural complexity of the understorey vegetation[101]. (5) Canopy gap fraction was estimated from hemispherical photographs at plot level as described in ref. [30]. Drone-based photogrammetry dated from September to October 2016 was used to further partition the canopy (in %) as (6) oil palm cover and (7) tree cover as described in ref. [102]. We also used the drone-based orthophotos to calculate (8) oil palm density as the number of living oil palms per plot irrespective of the orientation of the plot relative to the planting scheme (Supplementary Figs. 2 and 3) . For the smaller plots (5 × 5 m) unaffected by thinning, the oil palm density was simply the typical planting density in conventionally managed oil palm plantations (120 planted palms per hectare). Further details on the oil palm density calculation are given in the Supplementary Note 3. We also calculated (9) tree density as the number of trees planted and from natural regeneration per plot and expressed per hectare. We estimated the portion of the ground (in percent) as (10) understorey vegetation cover and (11) litter cover per subplot in February–March 2018. The understorey vegetation cover included all parts of plants lower than 130 cm in height, including the trunks and other parts of the planted trees but excluding oil palm trunks. (12) The litter depth (cm) was measured as the mean value in three randomly chosen positions inside each subplot with a metal ruler. To extract orthogonal axes (PC1 and PC2) that represent most of the variability in the vegetation structure, we applied a principal component analysis on all the structural variables after standardization to zero mean and unit variance.

Restoration outcomes

Ecosystem functioning

We measured 20 variables related to seven categories of ecosystem functioning including: productivity as (0) tree growth (basal area increment of the planted trees in $m^2\ ha^{-1}\ yr^{-1}$) that was further excluded from the analysis—see Supplementary Fig. 4, (1) oil palm yield (per island oil palm

yield changes in kg of fresh fruit bunches per island) and (2) above-ground biomass (biomass stored in the aerial parts of the planted trees and the oil palms, in t ha^{-1}); resistance to invasion as (3) native seeds (total number of arriving native seeds per m^2) and (4) resistance to invasive plants (100—observed cover of *Clidemia hirta*, in %); pollination as (5) pollinators (number of sampled individuals) and (6) pollination rate (fraction of flowers on phytometer plants that are pollinated, %); soil quality as (7) soil P (phosphorous content, %), (8) 1/soil C:N (the molar ratio of soil C to soil N concentration) and (9) soil decompaction (inverse of soil bulk density in g cm^{-3}); predation and herbivory as (10) predatory invertebrates (total activity duration of insectivorous bats and birds, in seconds); (11) predatory arthropods (number of sampled individuals), (12) predatory soil fauna (energy flux, in J h^{-1}), (13) herbivory (energy flux, in J h^{-1}); carbon and nutrient cycling as (14) decomposers (energy flux, in J h^{-1}); (15) litter decomposition (relative biomass loss of litter after 6 months in litterbags, %) and (16) litter input (biomass of leaf litter falling in traps, g m^{-2}); water and climate regulation as (17) evapotranspiration (canopy latent heat flux, in W m^{-2}); (18) soil water infiltration capacity (saturated soil hydraulic conductivity in cm h^{-1}) and (19) microclimate buffering (median daily amplitude of air temperature during 1 yr, °C d^{-1}). A more detailed summary of the 20 ecosystem functioning variables is presented in Supplementary Table 2.

Biodiversity

We derived taxonomic diversity for soil bacteria and soil fungi, soil fauna, herbs, trees, seeds, pollen, understorey arthropods, birds and bats. Most of the groups (arthropods, herbs, trees, birds and seeds) were sorted at the lowest possible taxonomic level (species or morphospecies). Pollen, soil fauna and bats were sorted to higher levels, mainly family, order and morphotypes, respectively. Soil bacteria and soil fungi were analysed by DNA-based marker gene sequencing as amplicons sequence variants or OTU, respectively. Hereafter, we refer to these different taxonomic units (species, family, order, morphotypes and OTU) as ‘species’ for simplicity.

Diversity was measured following the Hill number framework, which allows comparison across diversity indices that weigh relative abundances to varying extents (species richness, Shannon diversity and Simpson diversity) and are expressed in terms of effective numbers of species[103,104,105,106]. Species richness is more sensitive to locally rare species, Simpson diversity is more sensitive to locally dominant species and Shannon diversity favours neither rare nor dominant species. We show results for species richness and Simpson diversity in the main text and for all indicators in the Extended Data Tables 1 and 2 and Supplementary Tables 4– 8. The calculations were performed using the R packages iNext[82] and vegan[107].

Multidiversity and multifunctionality

Different indicators of biodiversity and ecosystem functioning were aggregated by calculating multidiversity and multifunctionality, respectively. Following ref. [18], we performed a cluster analysis to preselect indicators for achieving a representative measure of 'ecosystem function multifunctionality'. As tree growth and litter input were correlated and formed a cluster, we excluded tree growth from the analysis (Supplementary Note 4 and Supplementary Fig. 4). Following a threshold approach[108], we calculated multifunctionality (and multidiversity) as the number of ecosystem functioning (and biodiversity) indicators that cross a threshold, expressed as a certain percentage of the maximum observed values in our study landscape (among all 56 study plots). We calculated multifunctionality and multidiversity for all thresholds from 1% to 99% and presented results for a 50% threshold in the main text. To reduce the influence of extreme values, we used the mean of the three highest values observed in all study plots, respectively. As an alternative to the threshold approach, we also calculated multidiversity and multifunctionality as the average of the indicators[108]. Before multidiversity and multifunctionality calculations, all the variables were standardized to unit scale (for biodiversity and ecosystem functioning separately). The calculations were performed using the package multifunc in R[108].

Statistical analysis

Linear mixed-effect models

We used linear mixed-effect models to test the effects of the experimental treatment on restoration outcomes. We fitted three separated models for biodiversity using species richness, Shannon and Simpson diversity as response variables and one model for ecosystem functioning. These models included tree island (compared to our controls of conventionally managed oil palm monocultures), island area (plot edge length in m), planted diversity and the restoration outcome (either biodiversity or ecosystem functioning indicators) as single factors and tree island × indicator, island area × indicator, planted diversity × indicator, island area × planted diversity and island area × planted diversity × indicator interactions. For conventionally managed oil palm plots, island area was set to 10 m edge length and planted diversity to zero. Each response variable (biodiversity and ecosystem functioning indicators) was standardized to unit scale (between 0 and 1) as this improved the model diagnostics before applying the respective linear mixed-effect models; whereas we used logarithmic transformations for island area and planted diversity. Plot was included as a random term.

As an alternative to the linear mixed-effect models, we applied Kruskal–Wallis tests on each indicator of biodiversity and ecosystem functioning for comparison between the 52 tree islands and the four conventionally managed oil palm monocultures as control plots (Supplementary Note 5 and Supplementary Tables 8 and 9).

Structural equation modelling

We used piecewiseSEM[109] to assess the influence of tree island area and tree planted diversity on biodiversity and ecosystem functioning operating through increasing tree dominance, through differences in structural complexity (indirect effects) or through alternative mechanisms (direct effects). As a hypothetical causal model, we included direct paths between island area and tree dominance (PC2) and between tree planted diversity and structural complexity gradient (from open to dense and structurally complex vegetation, PC1; Extended Data Fig. 2). Piecewise SEMs are based on a set of linear equations which are evaluated individually[109]. For our analyses, we included:

$$lm(\text{restoration outcome} \approx \text{island area} + \text{planted diversity}) \quad (1)$$

$$lm(\text{structural complexity} \approx \text{planted diversity}) \quad (2)$$

$$lm(\text{tree dominance} \approx \text{island area}) \quad (3)$$

Across all restoration outcomes, the main variables were always included in the linear model (1). As tree dominance and structural complexity are potential mechanistic pathways explaining the influence of island area and tree planted diversity, alternative paths between them and biodiversity or ecosystem functioning were added, if they improved the model fit (based on modification indices, $P < 0.05$). Therefore, model selection influenced only the inclusion of structural complexity and tree dominance in the linear model (1). Effects of island area and planted diversity through mechanistic pathways were calculated by multiplying their effect on the mechanistic explanatory variable and the effect of the mechanistic explanatory variable on biodiversity or ecosystem functioning. Mechanisms that were not captured by either of our proposed mechanistic pathways are represented by the direct paths between island area and tree planted diversity and biodiversity or ecosystem functioning. We tested the assumption of normality of the residuals in models (1), (2) and (3) using Shapiro–Wilk normality test, applied a suitable transformation of the response variables if needed (package bestNormalize v.1.6.1). The transformation concerned four out of ten indicators for species richness and Shannon diversity, three indicators out of ten indicators for Simpson diversity and 14 indicators out of 19 for ecosystem functioning. Effect sizes were calculated using standardized coefficients. The island area and planted tree diversity were log-transformed as this improved the model fit. For each SEM, we quantified the goodness of fit using the following metrics: Fisher's *C* statistic and significance value based on a Chi-square test, the information criterion (Akaike information criterion (AIC), Bayesian information criterion (BIC), corrected AIC (AICc)) and pseudo-R^2 values and applied the test of directed separation as implemented in the package piecewiseSEM v.2.1.0[109].

Inclusion and ethics

The research included researchers from the Indonesian institutes Jambi University and Bogor Agricultural University throughout the research process—study design, study implementation, data ownership, intellectual property and authorship of publications. Local and regional research relevant to our study was considered in citations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw data are available at https://data.goettingen-research-online.de/dataverse/crc990, with the specific link for each dataset provided in Supplementary Tables 1–3. The processed data are available at https://doi.org/10.6084/m9.figshare.22320490. Seed DNA sequences are available in NCBI Genbank under the accession numbers OM811991–OM812021, OM837673–OM837724 and OM935782–OM935815. Sequencing data of the soil fungal community were deposited in the NCBI Sequence Read Archive (SRA) under Bioproject accession number PRJNA659225. The public UNITE database (https://unite.ut.ee/) v.7.2 on fungal ITS sequences was used as a reference of taxonomic classification. Sequence data of the bacterial communities were deposited in the NCBI SRA under Bioproject accession number PRJNA841353. Sequence identification was performed by mapping all curated sequences against the SILVA database v.132 (https://www.arb-silva.de/).

Code availability

The R code used in the current study is available at https://doi.org/10.6084/m9.figshare.22320490.

References

1. *UN Decade on Ecosystem Restoration* (UN, 2021); https://www.decadeonrestoration.org/about-un-decade.

2. Pashkevich, M. D. et al. Nine actions to successfully restore tropical agroecosystems. *Trends Ecol. Evol.* **37**, 963–975 (2022).

3. Lewis, S. L., Edwards, D. P. & Galbraith, D. Increasing human dominance of tropical forests. *Science* **349**, 827–832 (2015).

4. Meijaard, E. et al. The environmental impacts of palm oil in context. *Nat. Plants* **6**, 1418–1426 (2020).

5. Descals, A. et al. High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth Syst. Sci. Data* **13**, 1211–1231 (2021).

6. Qaim, M., Sibhatu, K. T., Siregar, H. & Grass, I. Environmental, economic, and social consequences of the oil palm boom. *Annu. Rev. Res. Econ.* **12**, 321–344 (2020).

7. Grass, I. et al. Trade-offs between multifunctionality and profit in tropical smallholder landscapes. *Nat. Commun.* **11**, 1186 (2020).

8. Dislich, C. et al. A review of the ecosystem functions in oil palm plantations, using forests as a reference system. *Biol. Rev.* **92**, 1539–1569 (2017).

9. *The IPBES Assessment Report on Land Degradation and Restoration* (IPBES, 2018).

10. Chazdon, R. & Brancalion, P. Restoring forests as a means to many ends. *Science* **365**, 24–25 (2019).

11. Brancalion, P. H. S. et al. Global restoration opportunities in tropical rainforest landscapes. *Sci. Adv.* **5**, eaav3223 (2019).

12. Coleman, E. A. et al. Limited effects of tree planting on forest canopy cover and rural livelihoods in Northern India. *Nat. Sustain.* **4**, 997–1004 (2021).

13. Kremen, C. & Merenlender, A. M. Landscapes that work for biodiversity and people. *Science* **362**, eaau6020 (2018).

14. Arroyo-Rodríguez, V. et al. Designing optimal human-modified landscapes for forest biodiversity conservation. *Ecol. Lett.* **23**, 1404–1420 (2020).

15. Corbin, J. D. & Holl, K. D. Applied nucleation as a forest restoration strategy. *For. Ecol. Manag.* **265**, 37–46 (2012).

16. Rey Benayas, J. M., Bullock, J. M. & Newton, A. C. Creating woodland islets to reconcile ecological restoration, conservation, and agricultural land use. *Front. Ecol. Environ.* **6**, 329–336 (2008).

17. Holl, K. D. et al. Applied nucleation facilitates tropical forest recovery: lessons learned from a 15-year study. *J. Appl. Ecol.* **57**, 2316–2328 (2020).

18. Manning, P. et al. Redefining ecosystem multifunctionality. *Nat. Ecol. Evol.* **2**, 427–436 (2018).

19. Teuscher, M. et al. Experimental biodiversity enrichment in oil-palm-dominated landscapes in Indonesia. *Front. Plant Sci.* **7**, 1538 (2016).

20. Wilson, E. O. & MacArthur, R. H. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).

21. Scherber, C. et al. Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature* **468**, 553–556 (2010).

22. Barry, K. E. et al. The future of complementarity: disentangling causes from consequences. *Trends Ecol. Evol.* **34**, 167–180 (2019).

23. Coverdale, T. C. & Davies, A. B. Unravelling the relationship between plant diversity and vegetation structural complexity: a review and theoretical framework. *J. Ecol*. https://doi.org/10.1111/1365-2745.14068 (2023).

24. Poorter, L. et al. Multidimensional tropical forest recovery. *Science* **374**, 1370–1376 (2021).

25. Iddris, N. A.-A. et al. Mechanical weeding enhances ecosystem multifunctionality and profit in industrial oil palm. *Nat. Sustain.* https://doi.org/10.1038/s41893-023-01076-x (2023).

26. Luke, S. H. et al. Riparian buffers in tropical agriculture: scientific support, effectiveness and directions for policy. *J. Appl. Ecol.* **56**, 85–92 (2019).

27. Bicknell, J. E. et al. Enhancing the ecological value of oil palm agriculture through set-asides. *Nat. Sustain.* https://doi.org/10.1038/s41893-022-01049-6 (2023).

28. Tews, J. et al. Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *J. Biogeogr.* **31**, 79–92 (2004).

29. Chase, J. M., Blowes, S. A., Knight, T. M., Gerstner, K. & May, F. Ecosystem decay exacerbates biodiversity loss with habitat loss. *Nature* **584**, 238–243 (2020).

30. Zemp, D. C. et al. Mixed-species tree plantings enhance structural complexity in oil palm plantations. *Agric. Ecosyst. Environ.* **283**, 106564 (2019).

31. Schall, P. et al. The impact of even-aged and uneven-aged forest management on regional biodiversity of multiple taxa in European beech forests. *J. Appl. Ecol.* **55**, 267–278 (2018).

32. Montoya-Sánchez, V. et al. Landscape heterogeneity and soil biota are central to multi-taxa diversity for landscape restoration. Preprint at *bioRxiv* https://doi.org/10.1101/2022.10.31.514517 (2022).

33. Gérard, A. et al. Oil-palm yields in diversified plantations: initial results from a biodiversity enrichment experiment in Sumatra, Indonesia. *Agric. Ecosyst. Environ.* **240**, 253–260 (2017).

34. Rahmani, T. A. et al. Evaluating the feasibility of oil palm agroforestry in Harapan Rainforest, Jambi, Indonesia. *For. Soc.* **5**, 458–477 (2021).

35. Zemp, D. C. et al. Tree performance in a biodiversity enrichment experiment in an oil palm landscape. *J. Appl. Ecol.* **56**, 2340–2352 (2019).

36. Darras, K. F. A. et al. Reducing fertilizer and avoiding herbicides in oil palm plantations—ecological and economic valuations. *Front. For. Glob. Change* https://doi.org/10.3389/ffgc.2019.00065 (2019).

37. Luke, S. H. et al. Managing oil palm plantations more sustainably: large-scale experiments within the Biodiversity and Ecosystem Function in Tropical Agriculture (BEFTA) Programme. *Front. For. Glob. Change* https://doi.org/10.3389/ffgc.2019.00075 (2020).

38. Azhar, B. et al. Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: evidence from Peninsular Malaysia. *Glob. Ecol. Conserv.* **3**, 553–561 (2015).

39. Watson, J. E. M. et al. The exceptional value of intact forest ecosystems. *Nat. Ecol. Evol.* **2**, 599–610 (2018).

40. Ahirwal, J., Sahoo, U. K., Thangjam, U. & Thong, P. Oil palm agroforestry enhances crop yield and ecosystem carbon stock in northeast India: implications for the United Nations sustainable development goals. *Sustain. Prod. Consum.* **30**, 478–487 (2022).

41. de Carvalho, W. R., Vasconcelos, S. S., Kato, O. R., Capela, C. J. B. & Castellani, D. C. Short-term changes in the soil carbon stocks of young oil palm-based agroforestry systems in the eastern Amazon. *Agrofor. Syst.* **88**, 357–368 (2014).

42. Khasanah, N. et al. Oil palm agroforestry can achieve economic and environmental gains as indicated by multifunctional land equivalent ratios. *Front. Sustain. Food Syst.* https://doi.org/10.3389/fsufs.2019.00122 (2020).

43. Oehri, J., Schmid, B., Schaepman-Strub, G. & Niklaus, P. A. Terrestrial land-cover type richness is positively linked to landscape-level functioning. *Nat. Commun.* **11**, 154 (2020).

44. Fahrig, L. et al. Resolving the SLOSS dilemma for biodiversity conservation: a research agenda. *Biol. Rev.* **97**, 99–114 (2022).

45. Paquette, A. et al. A million and more trees for science. *Nat. Ecol. Evol.* **2**, 763–766 (2018).

46. Drescher, J. et al. Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Phil. Trans. R. Soc. B* **371**, 20150275 (2016).

47. Bell, T. et al. A linear model method for biodiversity—ecosystem functioning experiments. *Am. Nat.* **174**, 836–849 (2009).

48. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).

49. Darras, K. F. A. et al. Assembling cheap, high-performance microphones for recording terrestrial wildlife: the Sonitor system. *F1000Research* **7**, 1984 (2021).

50. Darras K. F. A., Pérez N., M., Dilong L., Hanf-Dressler T., Markolf M., Wanger T. C. ecoSound-web: an open-source, online platform for ecoacoustics [version 2; peer review: 2 approved]. F1000Research, 9:1224 (2023) (https://doi.org/10.12688/f1000research.26369.2)

51. Darras, K., Furnas, B., Fitriawan, I., Mulyani, Y. & Tscharntke, T. Estimating bird detection distances in sound recordings for standardizing detection ranges and distance sampling. *Methods Ecol. Evol.* **9**, 1928–1938 (2018).

52. del Hoyo, J., Elliott, A., Sargatal, D., Christie, D. & de Juana, E. *Handbook of the Birds of the World Alive* (Lynx Editions, 2015).

53. Westphal, C. et al. Measuring bee diversity in different European habitats and biogeographical regions. *Ecol. Monogr.* **78**, 653–671

(2008).

54. van der Vecht, J. The Vespinae of the Indo-Malayan and Papuan areas (Hymenoptera, Vespidae). *Zool. Verh.* **34**, 1–82 (1957).

55. Bohart, R. M. & Menke, A. S. *Sphecid Wasps of the World: A Generic Revision* (Univ. California Press, 1976).

56. Yamane, S. A revision of the Japanese Eumenidae (Hymenoptera, Vespoidea) *Insecta matsumurana*. *J. Res. Fac. Agric. Hokkaido Univ.* **43**, 1–189 (1990).

57. Goulet, H. & Huber, J. T. *Hymenoptera of the World: An Identification Guide to Families* (Agriculture Canada, 1993).

58. Carpenter, J. & Nguyen, L. Keys to the genera of social wasps of South-East Asia (Hymenoptera: Vespidae). *Entomol. Sci.* **6**, 183–192 (2003).

59. Choate, P. M. *Key to the sub-Orders of Hymenoptera* (Univ. Florida, 2011); https://entnemdept.ufl.edu/choate/hymenoptera.pdf.

60. Engel, M. S. The honey bees of Indonesia (Hymenoptera: Apidae). *TREUBIA* **39**, 41–49 (2012).

61. Kempson, D., Lloyd, M. & Ghelardi, R. A new extractor for woodland litter. *Pedobiologia* **3**, 1–21 (1963).

62. Klarner, B. et al. Trophic niches, diversity and community composition of invertebrate top predators (Chilopoda) as affected by conversion of tropical lowland rainforest in Sumatra (Indonesia). *PLoS ONE* **12**, e0180915 (2017).

63. Potapov, A. M., Klarner, B., Sandmann, D., Widyastuti, R. & Scheu, S. Linking size spectrum, energy flux and trophic multifunctionality in soil food webs of tropical land-use systems. *J. Anim. Ecol.* **88**, 1845–1859 (2019).

64. Ehnes, R. B., Rall, B. C. & Brose, U. Phylogenetic grouping, curvature and metabolic scaling in terrestrial invertebrates. *Ecol. Lett.* **14**, 993–1000 (2011).

65. Ballauff, J. et al. Legacy effects overshadow tree diversity effects on soil fungal communities in oil palm-enrichment plantations. *Microorganisms* **8**, 1577 (2020).

66. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

67. Kõljalg, U. et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271–5277 (2013).

68. Berkelmann, D., Schneider, D., Hennings, N., Meryandini, A. & Daniel, R. Soil bacterial community structures in relation to different oil palm management practices. *Sci. Data* **7**, 421 (2020).

69. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

70. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

71. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

72. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

73. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

74. Chen, L. et al. GMPR: a robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* **6**, e4600 (2018).

75. White, T. J., Bruns, T., Lee, S. & Taylor, J. W. in *PCR Protocols: A Guide to Methods and Applications* (eds Innis, M. A. et al.) 315–322 (Academic Press, 1990).

76. Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105–1109 (1991).

77. Kress, W. J. & Erickson, D. L. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS ONE* **2**, e508 (2007).

78. CBOL Plant Working Group et al. A DNA barcode for land plants. *Proc. Natl Acad. Sci. USA* **106**, 12794–12797 (2009).

79. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).

80. Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S. & Kreft, H. Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biol. Conserv.* **213**, 234–242 (2017).

81. van Kleunen, M. et al. The Global Naturalized Alien Flora (GloNAF) database. *Ecology* **100**, e02542 (2019).

82. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).

83. Jantz, N., Homeier, J., León-Yánez, S., Moscoso, A. & Behling, H. Trapping pollen in the tropics—comparing modern pollen rain spectra of different pollen traps and surface samples across Andean vegetation zones. *Rev. Palaeobot. Palynol.* **193**, 57–69 (2013).

84. Stockmarr, J. Tablets with spores used in absolute pollen analysis. *Pollen Spores* **13**, 615–621 (1971).

85. Erdtman, G. *Handbook of Palynolgy: Morphology, Taxonomy, Ecology. An Introduction to the Study of Pollen Grains and Spores* (København

Munksgaard, 1969).

86. Jantz, N., Homeier, J. & Behling, H. Representativeness of tree diversity in the modern pollen rain of Andean montane forests. *J. Veg. Sci.* **25**, 481–490 (2014).

87. Pouliot, M., Bayala, J. & Ræbild, A. Testing the shade tolerance of selected crops under *Parkia biglobosa* (Jacq.) Benth. in an agroforestry parkland in Burkina Faso, West Africa. *Agrofor. Syst.* **85**, 477–488 (2012).

88. Prabowo, W. E. et al. Bird responses to lowland rainforest conversion in Sumatran smallholder landscapes, Indonesia. *PLoS ONE* **11**, e0154876 (2016).

89. Roldán Serrano, A. & Guerra-Sanz, J. M. Quality fruit improvement in sweet pepper culture by bumblebee pollination. *Sci. Hort.* **110**, 160–166 (2006).

90. Chave, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Change Biol.* **20**, 3177–3190 (2014).

91. Asari, N., Suratman, M. N., Jaafar, J. & Khalid, M. Md. Estimation of above ground biomass for oil palm plantations using allometric equations. *4th Int. Conf. Biol. Environ. Chem.* **58**, 110–114 (2013).

92. Zanne, A. E. et al. Data from: Towards a worldwide wood economics spectrum. *Dryad* https://doi.org/10.5061/DRYAD.234 (2009).

93. Krashevska, V. et al. Micro-decomposer communities and decomposition processes in tropical lowlands as affected by land use and litter type. *Oecologia* **187**, 255–266 (2018).

94. Ellsäßer, F. et al. Predicting tree sap flux and stomatal conductance from drone-recorded surface temperatures in a mixed agroforestry system—a machine learning approach. *Remote Sens.* **12**, 4070 (2020).

95. Ellsäßer, F., Röll, A., Stiegler, C., Hendrayanto, & Hölscher, D. Introducing QWaterModel, a QGIS plugin for predicting evapotranspiration from land surface temperatures. *Environ. Model. Softw.* **130**, 104739 (2020).

96. Timmermans, W. J., Kustas, W. P. & Andreu, A. Utility of an automated thermal-based approach for monitoring evapotranspiration. *Acta Geophys.* **63**, 1571–1608 (2015).

97. Donfack, L. S. et al. Microclimate and land surface temperature in a biodiversity enriched oil palm plantation. *For. Ecol. Manag.* **497**, 119480 (2021).

98. Isles, P. D. F. The misuse of ratios in ecological stoichiometry. *Ecology* **101**, e03153 (2020).

99. Ehbrecht, M., Schall, P., Ammer, C. & Seidel, D. Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate. *Agric. For. Meteorol.* **242**, 1–9 (2017).

100. Ehbrecht, M., Schall, P., Juchheim, J., Ammer, C. & Seidel, D. Effective number of layers: a new measure for quantifying three-dimensional stand structure based on sampling with terrestrial LiDAR. *For. Ecol. Manag.* **380**, 212–223 (2016).

101. Willim, K. et al. Assessing understory complexity in beech-dominated forests (*Fagus sylvatica* L.) in Central Europe—from managed to primary forests. *Sensors* **19**, 1684 (2019).

102. Khokthong, W. et al. Drone-based assessment of canopy cover for analyzing tree mortality in an oil palm agroforest. *Front. For. Glob. Change* https://doi.org/10.3389/ffgc.2019.00012 (2019).

103. Roswell, M., Dushoff, J. & Winfree, R. A conceptual guide to measuring species diversity. *Oikos* **130**, 321–338 (2021).

104. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies.

Ecol. Monogr. **84**, 45–67 (2014).

105. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).

106. Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427–432 (1973).

107. Oksanen, J. et al. vegan: Community ecology package. R version 2.5-6 https://cran.r-project.org/web/packages/vegan/index.html (2018).

108. Byrnes, J. E. K. et al. Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods Ecol. Evol.* **5**, 111–124 (2014).

109. Lefcheck, J. S. PiecewiseSEM: piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).

Acknowledgements

We thank PT Humusindo for granting us access to and use of their properties, as well as A. F. C. Rodríguez, A. Angulo, L. S. Donfack, M. Ehbrecht and D. Seidel for their contribution to data; the field assistants E. J. Siahaan, K. H. Dalimunthe and Juliandi for plot maintenance and for their technical assistance in data collection; and M. Teuscher and A. Gérard for their contribution to establishing the experiment. Illustrations were provided by F. Arndt, Formenorm Leipzig. This study was conducted using the research permit nos. 337/SIP/FRP/E5/Dit.KI/IX/2016 (D.C.Z.), 58/SIP/IV/FR/10/2021 and 20/SIP/IV/FR/1/2022 (C.M.), 53/EXT/SIP/FRP/E5/Dit.KI/IX/2017 (V.K.), 2C13UA0011-S (H.L.), 54/EXT/SIP/FRP/E5/Dit.KI/IX/2017 (A. Potapov), 4233/FRP/E5/Dit.KI/XII/2017 (L. Sachsenmaier), 343/SIP/FRP/E5/Dit.KI/X/2016 (A.K.), SK.336/KSDAE/SET/KSA.2/7/2019 (S.B. and C.A.S.), 1837/FRP/E5/Dit.KI/VIII/2017 (K.K.) and 1092/FRP/SM/VIII/2015 and 1487/FRP/SM/KI VI/2016 (W.K.) from the Indonesia Ministry of Research and Technology. The experiment used in this study is part of the global

network of tree diversity experiments TreeDivNet (https://treedivnet.ugent.be/). Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project ID 192626868—SFB 990 and the Ministry of Research, Technology and Higher Education (Ristekdikti) in the framework of the collaborative German–Indonesian research project CRC990 (D.C.Z., F.B., K.D., I.G., A. Potapov, A.R., I.A., J.B., H.B., D. Berkelmann, S.B., D. Buchori, R.D., O.G., F.E., R.F., N.H., B.I., W.K., V.K., A.K., J.K., K.L., H.L., M.M., M.S.M., C.C.M.M., Y.A.M., G.B.P., H.D.P., A. Polle, D.A.P., L. Sachsenmaier, S.S., F.S., A.C.S., L. Sundawati, T.T., M.W., D.H. and H.K.). Dorothea Schlözer Postdoctoral Programme of the Georg-August-Universität Göttingen (N.G.-R.).

Funding

Open access funding provided by University of Neuchâtel.

Author information

Authors and Affiliations

1. Conservation Biology, Institute of Biology, Faculty of Sciences, University of Neuchâtel, Neuchâtel, Switzerland

   Delphine Clara Zemp

2. Biodiversity, Macroecology and Biogeography, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Göttingen, Germany

   Delphine Clara Zemp, Nathaly Guerrero-Ramirez, Fabian Brambach, Gustavo B. Paterno, Lena Sachsenmaier & Holger Kreft

3. Centre of Biodiversity and Sustainable Land Use (CBL), University of Göttingen, Göttingen, Germany

   Delphine Clara Zemp, Oliver Gailing, Riko Fardiansah, Andrea Polle, Stefan Scheu, Meike Wollni, Dirk Hölscher & Holger Kreft

4. Agroecology, Department of Crop Sciences, Faculty of Agricultural Science, University of Göttingen, Göttingen, Germany

Kevin Darras, Isabelle Arimond, Kevin Li & Teja Tscharntke

5. Ecology of Tropical Agricultural Systems, Institute of Agricultural Sciences in the Tropics, University of Hohenheim, Stuttgart, Germany

Ingo Grass

6. Animal Ecology, J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Göttingen, Germany

Anton Potapov, Valentyna Krashevska, Alena Krause, Mark Maraun, Di Ajeng Prameswari & Stefan Scheu

7. Tropical Silviculture and Forest Ecology, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Göttingen, Germany

Alexander Röll, Florian Ellsäßer, Watit Khokthong, Johanna Kückes, Hendrik Lorenz & Dirk Hölscher

8. Functional Agrobiodiversity, Dept. of Crop Sciences, Faculty of Agricultural Science, University of Göttingen, Göttingen, Germany

Isabelle Arimond

9. Forest Botany and Tree Physiology, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Göttingen, Germany

Johannes Ballauff & Andrea Polle

10. Department of Palynology and Climate Dynamics, Albrecht-von-Haller-Institute for Plant Sciences, Göttingen, Germany

Hermann Behling & Christina Ani Setyaningsih

11. Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, University of Göttingen, Göttingen,

Germany

Dirk Berkelmann, Rolf Daniel & Dominik Schneider

12. Department of Palynology and Climate Dynamics, Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen, Göttingen, Germany

Siria Biagioni

13. Department of Plant Protection, Faculty of Agriculture, Institut Pertanian Bogor. Jl. Meranti, IPB Dramaga Campus, Bogor, Indonesia

Damayanti Buchori

14. Center for Transdisciplinary and Sustainability Sciences, IPB University, Jalan Pajajaran, Indonesia

Damayanti Buchori

15. Centre for Ecosystem Modeling and Monitoring, Facultad de Ciencias, Universidad Mayor, Santiago, Chile

Dylan Craven

16. Forest Genetics and Forest Tree Breeding, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Göttingen, Germany

Oliver Gailing & Carina C. M. Moura

17. Department of Natural Resources, University of Twente, Enschede, Netherlands

Florian Ellsäßer

18. Zoological Museum, Center of Natural History, Universität Hamburg, Hamburg, Germany

Riko Fardiansah

19. Faculty of Forestry, University of Jambi Jln Raya Jambi, Jambi, Indonesia

Riko Fardiansah, Bambang Irawan & Herni Dwinta Pebrianti

20. Biogeochemistry of Agroecosystems, Faculty of Agricultural Science, University of Göttingen, Göttingen, Germany

Nina Hennings

21. Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

Watit Khokthong

22. Chairs of Statistics and Econometrics, Faculty of Business and Economics, University of Göttingen, Göttingen, Germany

Miryam Sarah Merk

23. Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University, Kampus IPB Darmaga, Bogor, Indonesia

Yeni A. Mulyani

24. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

Lena Sachsenmaier

25. Systematic Botany and Functional Biodiversity, Institute of Biology, Leipzig University, Leipzig, Germany

Lena Sachsenmaier

26. Department of Forest Management, Faculty of Forestry and Environment, IPB University, Kampus IPB Darmaga, Bogor, Indonesia

Fitta Setiajiati & Leti Sundawati

27. Environmental and Resource Economics, Department of Agricultural Economics and Rural Development, Faculty of Agricultural Sciences, University of Göttingen, Göttingen, Germany

Meike Wollni

Contributions

D.C.Z., N.G.-R., H.K. and D.H. conceived this work. D.C.Z. and N.G.-R. developed the methodology. D.C.Z., N.G.-R. and M.S.M. created the software. D.C.Z., N.G.-R. and H.L. undertook the formal analysis. D.C.Z., F.B., K.D., A. Potapov, I.A., J.B., D. Berkelmann, S.B., F.E., R.F., N.H., W.K., V.K., A.K., J.K., K.L., H.L., C.C.M.M., H.D.P., D.A.P., L. Sachsenmaier, F.S. and C.A.S. conducted the investigations. D.S. obtained resources. D.C.Z. undertook data curation. D.C.Z., N.G.-R. and H.L. were responsible for visualization. H.K., D.H., M.W., I.G., H.B., D. Buchori, R.D., O.G., M.M., Y.A.M., A. Polle, S.S., T.T., L. Sundawati and B.I. obtained funding. D.C.Z., L. Sundawati and B.I. conducted the project administration. H.K., D.H., I.G., H.B., D. Buchori, R.D., O.G., M.M., Y.A.M., A. Polle, S.S., T.T., L. Sundawati, B.I. and A.R. were responsible for supervision. D.C.Z. and N.G.-R. wrote the original draft. F.B., K.D., I.G., A.P., A.R., G.B.P., D.C., M.W., H.K. and D.H. reviewed and edited the final manuscript.

Corresponding author

Correspondence to Delphine Clara Zemp.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Bernhard Schmid, Robert Nasi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Interaction between the experimental treatment (island, area, planted diversity) and biodiversity indicator.

(**a**) Interaction island × indicator. (**b**) Interaction area × indicator. (**c**) Interaction planted diversity × indicator. Lines/Solid points are linear mixed-effect model fits using ggeffect: (a) centre for the error bars indicate marginal means (b-c) bands indicate 95% confidence intervals with the centre line indicating marginal means using ggeffect. Points are observed values (n = 56 study plots). Only significant interaction terms in the analysis of variance of the linear mixed-effect models are shown ($p < 0.05$).

Extended Data Fig. 2 Principal components of the vegetation structure in the tree islands.

Each square represents one of the 52 tree islands, which vary in planted tree diversity (colour) and area (size of the square). The first two components (PC1 and PC2) explain 32% and 21% of the total variance, respectively. Vegetation structure variables included (variable loadings PC1 and PC2): Oil palm cover (−0.13, 1.26), ENL, i.e., effective number of layers (−0.83,

0.65), understorey cover (−1.10, −0.07), canopy gap fraction (−1.12, −0.41), litter depth (−0.45, −0.79), tree cover (0.73, −0.99), litter cover (0.88, −0.55), MeanFRAC, i.e., a measure of the geometric complexity of the vegetation structure that is density-dependent (1.20, −0.10), tree density (0.64, 0.23), UCI, i.e., understorey complexity index (0.93, 0.35), SSCI, i.e., stand structural complexity index (0.85, 0.49) and oil palm density (0.64, 0.23). PC1 is mostly associated with dense and structurally complex vegetation and with low understorey vegetation cover (see Methods for details). PC2 is mostly associated with a high proportion of palms and low proportion of trees in the canopy. In the main text, we refer to PC1 as 'structural complexity' and minus PC2 as 'tree dominance'.

Extended Data Fig. 3 Direct and indirect influence of tree island size and planted tree diversity on multidiversity and multifunctionality.

Results of the structural equation models (i.e. standard coefficient estimates) for multifunctionality and multidiversity (based on species richness, Shannon diversity and Simpson diversity) with alternative calculation methods (average or thresholding approach - see Methods). The effect of tree island area can be direct (black bars) or *via* tree dominance (grey bars), and the effect of planted tree diversity can be direct (dark green bars - here absent) or via structural complexity (light green bars). Only significant estimates are presented ($p < 0.05$ two-sided ANOVA, $n = 52$ tree islands).

Extended Data Fig. 4 Changes in per area yield across four years for different planted diversity levels.

Relative changes in per area yield ($\Delta Y_{ha}/Y_{ha_ref}$ in %) compared to conventionally managed oil palm monocultures ('reference') for different planted tree species diversity: 0, 1, 2, 3, 6 species, and all 52 tree islands combined ("total"). Plus signs represent means, horizontal lines medians, boxes interquartile ranges and vertical bars ranges. The dashed horizontal line is the mean of the reference palms (i.e. the palms in adjacent position 3). 'n'-values denote the number of months included in the annual boxes. Significant differences to reference as based on Mann-Whitney tests are

indicated at the bottom of each panel, with the significance levels * ($p < 0.1$), ** ($p < 0.05$) and *** ($p < 0.01$).

Extended Data Fig. 5 Effects of the experimental treatment on per palm and per area yield.

Each dot represents an individual oil palm that was monitored within our study (N = 404 in total, from October 2017 to November 2018). In each boxplot, the horizontal line corresponds to the median; the lower and upper hinges correspond to the first and third quartiles and the upper and lower whiskers are limited by the 1.5 the interquartile ranges. The overall effects of the experimental treatment on (**a**) per palm yield (in kg/palm) and (**b**) per area yield (in kg/ha) were significant in both cases ($\chi^2 = 6.39$, df = 2, p-value = 0.04 and $\chi^2 = 75.35$, df = 2, p-value < 0.001, respectively). Multi-comparison indicates that (**a**) per palm yield inside the islands was lower than per palm yield adjacent to the tree islands; (**b**) per area yield inside the tree islands were lower than per area yield adjacent to the tree islands and per area yield in the conventionally managed oil palm monocultures. **c**) There was no effect of the distance to the tree island edge (i.e. positions 1, 2 and 3) on the per palm yield of the adjacent oil palms ($\chi^2 = 4.47$, df = 2, p-value = 0.10). **d**) There was no effect of tree planting on the per palm yield of the adjacent oil palms ($\chi^2 = 1.02$, df = 1, p-value = 0.31). **e**) The effect of oil palm thinning on per palm yield of the adjacent palms was marginally significant ($\chi^2 = 3.78$, df = 1, p-value = 0.05). Multi-comparisons were conducted with post hoc Tukey tests. More details of the statistical analysis are provided in section 2.3.2 of[19], where models number one, two, three and four correspond here to the panels a), b), c) and e), respectively. Significance levels are indicated by. ($p < 0.1$), * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

Extended Data Table 1 ANOVA of the linear mixed-effect models for biodiversity and ecosystem functioning

Extended Data Table 2 Summary statistics of piecewise structural equation models for each biodiversity and ecosystem functioning indicators and multidiversity and multifunctionality

Extended Data Table 3 Result of the principal component analysis of the vegetation structure variables. (n = 52 tree islands)

Extended Data Table 4 List of soil fauna groups and associated guild

Supplementary information

Supplementary Information

Supplementary Notes 1–5, Figs. 1–4 and Tables 1–10.

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Zemp, D.C., Guerrero-Ramirez, N., Brambach, F. *et al.* Tree islands enhance biodiversity and functioning in oil palm landscapes. *Nature* **618**, 316–321

(2023). https://doi.org/10.1038/s41586-023-06086-5

- Received: 07 April 2022
- Accepted: 14 April 2023
- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06086-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

‘Tree islands’ give oil-palm plantation a biodiversity boost

- Benjamin Thompson
- Shamini Bundell

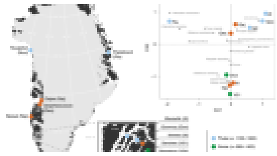
Nature Nature Podcast 24 May 2023

Tree islands boost biodiversity in oil-palm plantations

- Robert Nasi

Nature News & Views 24 May 2023

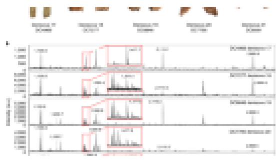
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06086-5



Ancient DNA provides insights into 4,000 years of resource economy across Greenland

06 October 2022

Frederik V. Seersholm, Hans Harmsen, … Anders J. Hansen



The earliest Denisovans and their cultural adaptation

25 November 2021

Samantha Brown, Diyendo Massilani, … Katerina Douka



Neolithic dental calculi provide evidence for environmental proxies and consumption of wild edible fruits and herbs in central Apennines

19 December 2022

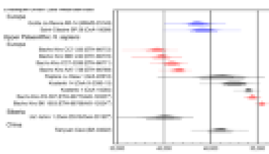
Alessia D’Agostino, Gabriele Di Marco, … Angelo Gismondi



[Palaeoproteomics gives new insight into early southern African pastoralism]

02 September 2020

Louise Le Meillour, Séverine Zirah, … Joséphine Lesur



[Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria]

11 May 2020

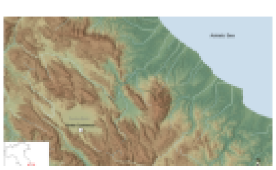
Jean-Jacques Hublin, Nikolay Sirakov, … Tsenka Tsanova



[Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza]

04 December 2018

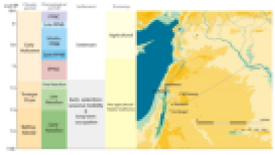
Pierre Zalloua, Catherine J. Collins, … Elizabeth Matisoo-Smith



Multipronged dental analyses reveal dietary differences in last foragers and first farmers at Grotta Continenza, central Italy (15,500–7000 BP)

19 February 2021

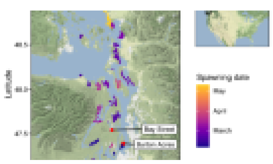
Alessia Nava, Elena Fiorin, … Emanuela Cristiani



Multi-isotope evidence of population aggregation in the Natufian and scant migration during the early Neolithic of the Southern Levant

04 June 2021

Jonathan Santana, Andrew Millard, … Zeidan Kafafi



Ancient DNA reveals phenological diversity of Coast Salish herring harvests over multiple centuries

06 August 2022

Eleni L. Petrou, Robert Kopperl, … Lorenz Hauser

- Article
- Open Access
- Published: 03 May 2023

Ancient human DNA recovered from a Palaeolithic pendant

- Elena Essel ,
- Elena I. Zavala,
- Ellen Schulz-Kornas,
- Maxim B. Kozlikin,
- Helen Fewlass,
- Benjamin Vernot,
- Michael V. Shunkov,
- Anatoly P. Derevianko,
- Katerina Douka,
- Ian Barnes,
- Marie-Cécile Soulier,
- Anna Schmidt,
- Merlin Szymanski,
- Tsenka Tsanova,
- Nikolay Sirakov,
- Elena Endarova,
- Shannon P. McPherron,
- Jean-Jacques Hublin,
- Janet Kelso,
- Svante Pääbo,
- Mateja Hajdinjak,
- Marie Soressi &
- …
- Matthias Meyer

Show authors

Nature **volume 618**, pages 328–332 (2023)

- 54k Accesses

- 1 Citations

- 1888 Altmetric
- Metrics details

Abstract

Artefacts made from stones, bones and teeth are fundamental to our understanding of human subsistence strategies, behaviour and culture in the Pleistocene. Although these resources are plentiful, it is impossible to associate artefacts to specific human individuals[1] who can be morphologically or genetically characterized, unless they are found within burials, which are rare in this time period. Thus, our ability to discern the societal roles of Pleistocene individuals based on their biological sex or genetic ancestry is limited[2,3,4,5]. Here we report the development of a non-destructive method for the gradual release of DNA trapped in ancient bone and tooth artefacts. Application of the method to an Upper Palaeolithic deer tooth pendant from Denisova Cave, Russia, resulted in the recovery of ancient human and deer mitochondrial genomes, which allowed us to estimate the age of the pendant at approximately 19,000–25,000 years. Nuclear DNA analysis identifies the presumed maker or wearer of the pendant as a female individual with strong genetic affinities to a group of Ancient North Eurasian individuals who lived around the same time but were previously found only further east in Siberia. Our work redefines how cultural and genetic records can be linked in prehistoric archaeology.

Main

Palaeolithic assemblages typically contain a multitude of objects that may differ in age by hundreds or thousands of years, even when found in close proximity[1]. Thus, it can be challenging to associate human remains with specific objects. Recent advances in the retrieval of human DNA from sediments[6,7,8] can be used to connect artefacts with genetic populations. However, precise identification of the specific makers or users of these objects would require the recovery of human DNA directly from the objects themselves, analogous to modern-day forensic investigations. In theory, such analyses are most promising for artefacts made from animal bones or teeth,

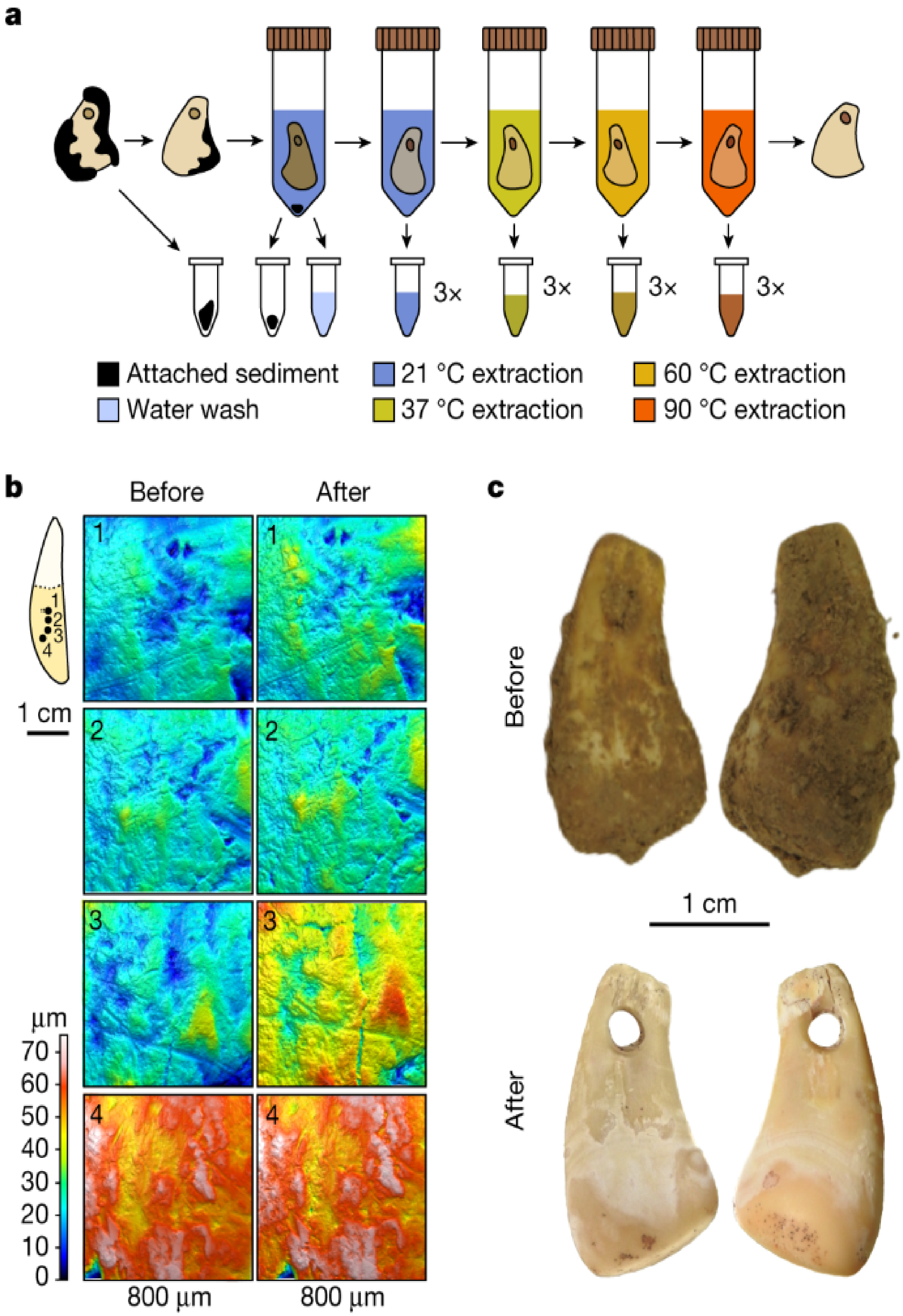
not only because they are porous and thereby conducive to the penetration of body fluids (for example, sweat, blood or saliva) but also because they contain hydroxyapatite, which is known to adsorb DNA and reduce its degradation by hydrolysis and nuclease activity[9,10]. Ancient bones and teeth may therefore function as a trap not only for DNA that is released within an organism during its lifetime and subsequent decomposition but also for exogenous DNA that enters the matrix post-mortem through microbial colonization[11] or handling by humans. However, DNA extraction from ancient skeletal material either requires destructive sampling, or risks alteration of specimens if they are directly submerged in extraction buffer[12,13]. Conservation is a primary concern because of the scarcity of bone and tooth artefacts at Pleistocene sites, especially of pendants and other ornaments that were extensively handled or worn in close body contact. We therefore set out to develop a method for DNA isolation from bones and teeth that preserves the integrity of the material, including surface microtopography, and to investigate the possibility of DNA retrieval from bone and tooth artefacts.

A non-destructive DNA isolation method

To identify reagents compatible with non-destructive DNA extraction, we selected ten unmodified faunal remains from the Palaeolithic sites of Quinçay and Les Cottés in France (Extended Data Table 1 and Supplementary Information 1), which were similar in size and shape to material typically used for osseous artefact production, and submerged them in several reagents previously used in ancient DNA extraction, as well as in water for comparison. These included (1) a guanidinium thiocyanate-containing reagent previously suggested for non-destructive DNA extraction[12], (2) an ethylenediaminetetraacetate (EDTA) solution, which is a decalcifier commonly used in ancient DNA extraction[14,15,16], (3) a sodium hypochlorite (bleach) solution, which is an oxidizing reagent used to remove surface-exposed contaminant DNA[11,17], and (4) a sodium phosphate buffer supplemented with detergent[11], which has been recently shown to enable temperature-controlled DNA release from powdered bone samples[18].

Mapping of the microtopography using quantitative 3D surface texture analysis[19,20] before and after the treatments revealed substantial surface alterations on all objects exposed to either the guanidinium thiocyanate reagent or EDTA (Extended Data Fig. 1 and Supplementary Information 2). By contrast, only sporadic and smaller alterations were detected with the other reagents, including sodium phosphate buffer (Fig. 1b), possibly due to the removal of traces of sediment and other small particles, as indicated by visible changes in coloration of some of the objects (Extended Data Fig. 2). On the basis of these results, we developed a non-destructive DNA isolation method for the stepwise release of DNA from the bone or tooth matrix using serial incubations in sodium phosphate buffer at 21, 37, 60 and 90 °C, with three incubations per temperature (Fig. 1a).

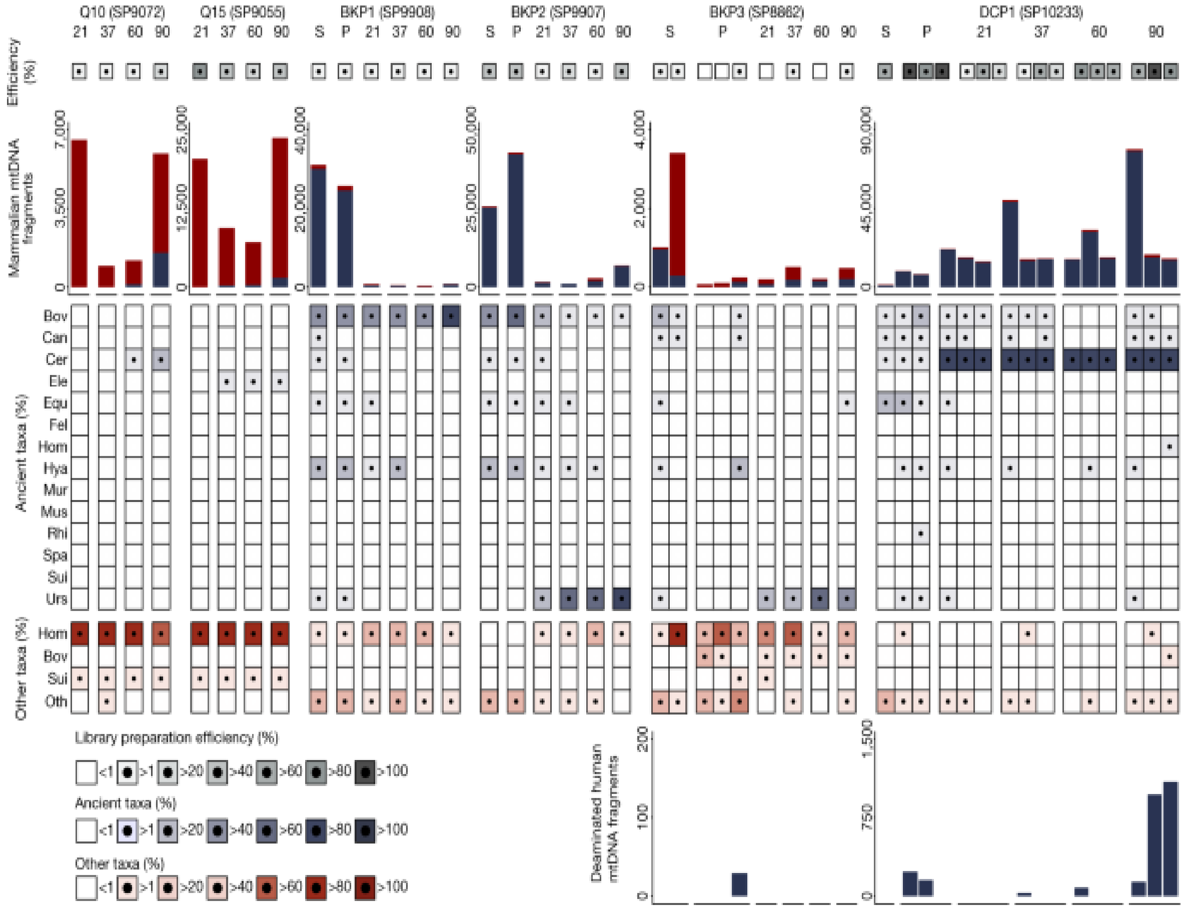
Fig. 1: Overview of the non-destructive DNA extraction method.


a
3×
3×
3×
3×
Attached sediment
Water wash
21 °C extraction
37 °C extraction
60 °C extraction
90 °C extraction
b
Before
After
1
2
3
4
1 cm
µm
70
60
50
40
30
20
10
0
800 µm
800 µm
c
Before
1 cm
After

a, Workflow of the gradual, non-destructive DNA extraction method using sodium phosphate buffer at elevated temperatures. **b**, Four 3D surface texture measurements (1–4) indicated on the outline of a tooth used for testing (SP6649) before and after non-destructive DNA extraction showing no substantial surface alterations. **c**, Photographs of DCP1 before and after cleaning and non-destructive DNA extraction.

We then applied this method to 11 osseous objects, labelled Q10 to Q19 as well as Q27, that were excavated several decades ago in the Châtelperronian layers of Quinçay Cave in France and that had potentially been used as tools some 35–45 thousand years ago (ka)[21] (Extended Data Table 1 and Extended Data Fig. 3). We prepared single-stranded DNA libraries[22,23] from the first DNA fraction recovered at each temperature, and enriched the libraries for mammalian mitochondrial (mt) DNA[24]. A metagenomic pipeline for assigning sequenced mtDNA fragments to mammalian taxa on the biological family level[6] identified 1,628 cervid mtDNA fragments in the 60 °C and 90 °C fractions of object Q10, a reindeer bone (109 and 1,519 fragments, respectively; Fig. 2 and Supplementary Data 1). These fragments showed elevated frequencies of cytosine (C)-to-thymine (T) substitutions at their ends, consistent with deamination of cytosine residues seen in ancient DNA[6,25] (Supplementary Data 1). Another object, Q15, which was made of ivory, yielded 2,004 elephantid mtDNA sequences with elevated frequencies of C-to-T substitutions in the 37 °C, 60 °C and 90 °C fractions (248, 325 and 1,431 fragments, respectively). In addition, we identified hominid and suid mtDNA fragments with no evidence for ancient DNA base damage in every DNA fraction from the 11 objects, thus resulting from contamination with human and pig DNA after excavation. Human DNA contamination was particularly severe, amounting to between 70.9% and 98.3% of the identified mtDNA fragments (between 293 and 92,949 fragments in total; 17,627 on average), thereby potentially masking traces of ancient human or other mammalian DNA.

Fig. 2: Number and taxonomic composition of mammalian and human mtDNA fragments recovered from six artefacts during stepwise, non-destructive DNA extraction.



DNA fractions are denoted S (attached sediment), P (sediment pellet recovered during water wash), and 21, 37, 60 and 90 (three incubations in phosphate buffer at the indicated temperatures in °C). Low library preparation efficiencies indicate reduced DNA recovery due to the co-extraction of inhibitory substances. Assignments to ‘ancient’ and ‘other’ taxa were performed independently for each family based on the significance of evidence for cytosine deamination. The bottom right chart shows the number of deaminated hominin mtDNA fragments recovered in each fraction after hominin-specific mtDNA capture (only positive fractions). Bov, Bovidae; Can, Canidae; Cer, Cervidae; Ele, Elephantidae; Equ, Equidae; Fel, Felidae; Hom, Hominidae; Hya, Hyaenidae; Oth, other; Mur, Muridae; Mus, Mustelidae; Rhi, Rhinocerotidae; Spa, Spalacidae; Sui, Suidae; Urs, Ursidae.

Studying freshly excavated artefacts

As present-day human DNA contamination seemed to be ubiquitous on surfaces of objects that were handled with bare hands during and after excavation, we collected artefacts from ongoing excavations at two Palaeolithic sites, using gloves and facemasks as soon as they became partly exposed to prevent contamination. At Bacho Kiro Cave in Bulgaria, we recovered three Upper Palaeolithic tooth pendants (henceforth 'BKP1–BKP3') from layers I, H/I and I/J of niche 1 (Extended Data Fig. 3). At Denisova Cave, a tooth pendant ('DCP1') was recovered from layer 11 of the south chamber (Fig. 1c and Extended Data Fig. 4).

Larger clumps of sediment adhering to the artefacts were manually removed, and the artefacts were subsequently cleaned by three successive water washes. DNA was then extracted from the sediment clumps, the water was used for washing and the sediment particles were collected in this process ('sediment pellets'), as well as directly from the artefacts using the non-destructive method described above. Ancient mammalian mtDNA was detected in all fractions that were analysed, except for some of the water washes and the associated sediment pellets (Fig. 2). The trajectories of DNA released from the four artefacts were similar in that the highest yield of mammalian mtDNA was obtained at 90 °C in the phosphate-based DNA extraction (up to 734, 6,614, 456 and 77,910 mtDNA fragments for BKP1, BKP2, BKP3 and DCP1, respectively; Supplementary Data 1). However, library preparation efficiencies were low for the Bacho Kiro Cave material (less than 10% for many fractions), presumably due to the co-extraction of inhibitory substances (Fig. 2), indicating that more DNA was released than could be recovered and sequenced.

The phosphate DNA fractions obtained at 37 °C, 60 °C and 90 °C are dominated by ancient ursid mtDNA fragments for BKP2 and BKP3, and cervid mtDNA fragments for DCP1, in agreement with their morphological identification (Extended Data Table 1). For BKP1, which is morphologically indeterminate, the phosphate fractions are dominated by bovid mtDNA fragments. In contrast to the phosphate fractions, DNA recovered from the sediment adhering to the artefacts is taxonomically more heterogeneous (Fig. 2). In addition, substantially lower numbers of human mtDNA fragments (between 0 and 2,969 per fraction, 246 on average; Supplementary Data 1) were recovered from the freshly excavated artefacts than from the Quinçay material. Similarly, very few suid mtDNA fragments

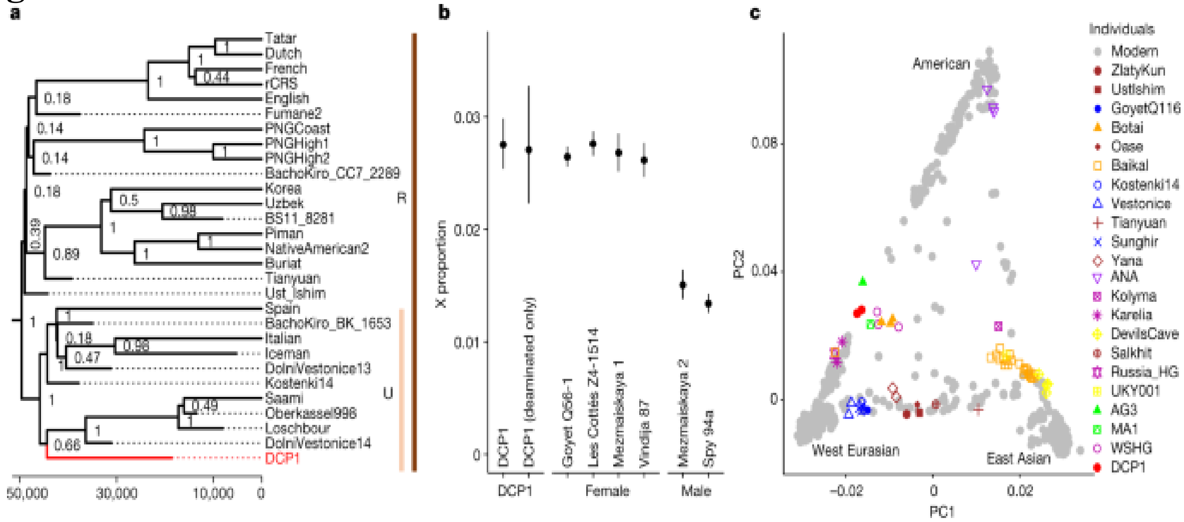
(15 or less) were recovered, indicating that little post-excavation contamination had occurred. Of note, significant signals of cytosine deamination were observed among the human mtDNA fragments recovered in one of the 90 °C fractions from DCP1.

To increase human DNA recovery, we enriched all libraries again using a capture probe set specifically targeting human mtDNA. For the Bacho Kiro Cave material, this enabled the detection of small traces of ancient human DNA in a sediment pellet from BKP3 (29 deaminated mtDNA fragments), but none of the other fractions. For DCP1, human mtDNA fragments with significant evidence for ancient DNA base damage were identified in the first two sediment pellets recovered from the water washes, the first 37 °C and 60 °C fractions, and all three 90 °C fractions (Fig. 2). The largest numbers of deaminated human mtDNA fragments were obtained in the second and third 90 °C fractions (971 and 1,096, respectively), indicating that extended incubation at high temperature enabled ancient human DNA release from the pendant.

Preparation of additional libraries from the second 90 °C fraction of DCP1, the fraction with the lowest estimate of present-day human contamination (0.1%, 95% CI: 0.0–2.8%), yielded 62-fold average coverage of the human mtDNA genome and a near-complete consensus sequence (Supplementary Information 5). This sequence, which falls together with mtDNA sequences assigned to haplogroup U in a phylogenetic tree (Fig. 3a), contains seven 'diagnostic' positions that distinguish it from the mtDNA sequences of other human individuals (Extended Data Table 2). Among the mtDNA fragments overlapping these positions, 86.6% (95% CI: 82.2–90.5%) match the state of DCP1 (Extended Data Table 3), suggesting that the mtDNA fragments recovered in this fraction originate predominantly, but not exclusively, from a single ancient human individual, presumably the user or the maker of the pendant. Support for the DCP1 consensus sequence is slightly lower in the first (77.8%, 95% CI: 40.0–97.2%) and third (82.8%, 95% CI: 78.9–86.2%) 90 °C phosphate fractions, consistent with the slightly higher estimates of present-day human contamination in these fractions (12.8%, 95% CI: 1.0–24.6% and 6.6%, 95% CI: 4.3–8.9%, respectively). By contrast, support for the DCP1 consensus is low in the preceding 60 °C fraction (37.5%, 95% CI: 8.5–75.5%), indeterminate in the 37 °C fraction and low in the first (20.0%, 95% CI: 5.7–43.7%) and second (9.5%, 95% CI: 1.2–30.4%) sediment

pellets. These results indicate that the initial water washes and incubations in phosphate buffer at below 90 °C primarily released ancient human DNA from one or more other individuals, which were present in smaller quantities in the surrounding sediment or had adsorbed directly to the surface of DCP1.

Fig. 3: Ancient human mtDNA and nuclear DNA isolated from DCP1.



a, The position of DCP1 in a Bayesian tree reconstructed from modern[31,32] and ancient[33] human mtDNA sequences (see Supplementary Information 5 for the full tree). Nodes are labelled with the corresponding posterior probabilities, and the *x* axis represents years from the present. Identified haplogroups are outlined by the bars on the right. rCRS, revised Cambridge reference sequence. **b**, X–autosome proportion in DCP1 (using all and deaminated molecules only) in comparison to data from six other ancient hominin individuals[34,35]. Circles correspond to the calculated values of the ratios for the number of X to (X + autosomal) fragments for each individual (*n* (of single-nucleotide polymorphisms (SNPs)) = 20,526, 3,734, 124,862, 85,901, 34,756, 41,632, 34,677 and 72,992 SNPs for each calculation, as ordered on the *x* axis). The error bars represent 95% binomial CIs of the measurement in each individual. **c**, Principal component (PC) analysis of non-African modern human genomes[36] (grey) with ancient human genomes (coloured) projected on top. DCP1 was analysed twice, using all data versus deaminated fragments only. AG3, Afontova Gora 3; ANA, ancient Native Americans; MA1, Mal’ta 1; Russia_HG, Russian hunter–gatherers; UKY001, Ust Kyakhta; WSHG, West Siberian hunter–gatherers.

Phylogenetic analyses and dating

On the basis of the branch length of the DCP1 consensus sequence in a tree with other present-day and ancient human mtDNA genomes (Fig. 3a), we estimated its age to 18.5 thousand years (kyr), with a 95% highest posterior density interval ranging from 4.6 to 31.6 kyr (Supplementary Information 5). Furthermore, we used cervid mtDNA probes to reconstruct the complete mtDNA genome of the tooth, identified as wapiti (*Cervus canadensis*), at 635-fold coverage. A tree with eight additional ancient wapiti mtDNA genomes of known age generated in this study (Supplementary Information 6), estimates the age of DCP1 at 24.7 kyr (highest posterior density interval of 12.8–39.0 kyr). Both genetic ages are consistent with each other and with the younger of two radiocarbon dates that we obtained from charcoal discovered in the proximity of DCP1 in layer 11 (OxA-X-3089-11: 24,200–23,830 calibrated years before present (cal bp) and OxA-X-3089-12: 39,180–37,560 cal bp) at 95.4% probability (Supplementary Information 1). We therefore suggest that genetic dating obviates the need for direct radiocarbon dating of the pendant, although this remains technically possible after non-destructive DNA extraction (Supplementary Information 3).

For nuclear DNA analysis, hybridization capture was performed using libraries from the second and third 90 °C phosphate fractions, targeting sites in the human genome that are known to be polymorphic in modern or archaic humans and that are located in regions of high sequence divergence between humans and other mammals[8]. Sequence information was obtained for 336,429 of these sites (71.5% of the sites targeted), with estimates of present-day human and faunal contamination both below 1%. Comparisons with present-day human populations[26] using *f3*-statistics and *D*-statistics[27,28] show high affinities to Native Americans (Extended Data Fig. 5). When projected into a principal component analysis with other ancient human individuals (Fig. 3c), DCP1 falls within a group of Ancient North Eurasian individuals from further east in Siberia, which includes the approximately 24 ka Mal'ta 1 and the approximately 17 ka Afontova Gora 3 individuals[29,30]. Both of these individuals are genetically closer to DCP1 than non-Ancient North Eurasian individuals when tested with *D*-statistics (Extended Data Fig. 6b), and all three show similar affinities to ancient Siberians and Native Americans with *f3*-statistics and *D*-statistics (Extended

Data Fig. 6a,c). In addition, shotgun data were produced from one of the libraries to allow a comparison of sequence coverage for the X chromosome and the autosomes, which is compatible with the human DNA in the 90 °C fraction originating predominantly from a female individual (Fig. 3b and Supplementary Information 7).

Conclusions

In summary, our work highlights that artefacts made from bones or teeth are a previously untapped source of ancient human DNA that can provide insights about the ancestry and biological sex of the individuals who handled, carried or wore these objects in the deep past. The non-destructive DNA extraction method reported here allows a stepwise release of this DNA, making it possible to distinguish DNA that penetrated deeply into an object during its manufacture or use from DNA that may originate from the surrounding sediment. Of note, the coverage depth of targeted sites in the human nuclear genome achieved from DCP1 is similar to what has been obtained with hybridization capture from well-preserved Pleistocene human remains[30]. Furthermore, the recovery of both human and faunal DNA enabled two independent genetic estimates of its age.

Further work is needed to determine how often human DNA can be recovered from Palaeolithic osseous artefacts. As surface DNA contamination can hamper these analyses, we urge archaeologists to apply protocols for minimizing handling during and after excavation. If this is done, it might become possible to systematically combine genetic and cultural analyses to study Pleistocene artefact use and uncover possible task specialization by individuals of a particular biological sex or genetic ancestry.

Methods

Sample collection

For reagent testing, we collected ten unmodified faunal remains, similar in and size and shape to material used in artefact production, from the late

Middle Palaeolithic and early Upper Palaeolithic deposits of the French sites Quinçay (seven) and Les Cottés (three) (Extended Data Table 1 and Extended Data Fig. 2). The estimated ages of these specimens range from 55 to 35 kyr[21,37,38]. Non-destructive DNA extraction was then applied to 15 osseous specimens excavated at Quinçay Cave, all from layers attributed to the Châtelperronian technocomplex and probably dating to 45–35 ka[21,37], to three tooth pendants excavated in the Initial Upper Palaeolithic (45–43 ka)[39] layers in the niche 1 area of Bacho Kiro Cave[40,41], as well as on one tooth pendant excavated in 2019 in layer 11 (39–24 ka), square E-3, of the south chamber of Denisova Cave (Extended Data Table 1 and Extended Data Fig. 3). Samples from Bacho Kiro and Denisova Cave were excavated and handled using sterile gloves and additional precautions were taken to minimize the introduction of modern DNA contamination. More information about the samples and the archaeological context in which they were recovered is provided in Supplementary Information 1.

Testing reagents for non-destructive DNA extraction using 3D surface texture measurements

We evaluated four reagents for their potential use in non-destructive DNA extraction by applying each to two Pleistocene samples comparable with the ones usually transformed into bone or tooth tools and ornaments, preferably choosing one bone and one tooth fragment for each reagent (Extended Data Table 1). As none of the objects was perfectly clean and sediment microparticles may be released upon exposure to liquids, we also performed treatments with water to obtain baseline measurements of changes in the microtopography that are independent of the chemical compositions of the reagents. Reagents and incubation conditions were as follows:

1. (1)

   Following a protocol[42] further detailing the non-destructive DNA extraction method for museum specimens by Rohland et al. (2004)[12], samples were completely submerged in 10–15 ml guanidinium thiocyanate buffer (5 M guanidinium isothiocyanate, 50 mM Tris-HCl, pH 8.0, 25 mM NaCl, 1.3% Triton X-100, 20 mM EDTA, pH 8.0, and 50 mM DTT) and incubated for 5 days in the dark.

2. (2)

   Incubation in EDTA, supplemented with a detergent (0.45 M EDTA, pH 8.0, and 0.05% Tween-20), was performed for 15 min at room temperature. EDTA is widely used in bone extraction protocols as a decalcifying agent to dissolve hydroxyapatite, the main mineral component of bones[14,15,16]. It is therefore expected to cause severe alterations to the sample material and was included to demonstrate the effect of such alterations on the quantitative 3D-surface texture analysis (3DST) measurements.

3. (3)

   Two samples were submerged for 15 min in 0.5% sodium hypochlorite (bleach) solution, a reagent previously used for decontamination of ancient skeletal remains[11,17]. Bleach treatment destroys surface-bound DNA and is not suitable for DNA extraction. However, it may be used in later implementations of the method to remove contaminant DNA before non-destructive DNA extraction.

4. (4)

   Temperature-controlled release of DNA was performed following a modified version of the method by Essel et al. (2021)[18] by submerging the samples in sodium phosphate buffer (0.5 M sodium phosphate, pH 7.0, and 0.1% Tween-20). Serial incubations were performed at 21 (room temperature), 37, 60 and 90 °C, each with incubation times of 30 min for a total of three incubations per temperature.

All treatments were performed in 50-ml Falcon tubes without agitation to avoid any mechanical damage to the sample. Reagent volumes were chosen individually for each sample (ranging from 5 ml to 17.5 ml) to ensure complete submergence. Incubations above room temperature were performed in a Heating-ThermoMixer MHR 11 (Hettich Benelux) equipped with inserts for 50-ml Falcon tubes. For the 90 °C incubation in phosphate buffer, the temperature of the device was set to 99 °C to ensure that 90 °C was reached inside the tube by the end of the incubation time. After the treatments, all samples were placed in fresh 50-ml Falcon tubes and

incubated in water for 1 h at room temperature to remove residual reagents. Samples were dried at room temperature for 5 days before they were returned to their storage containers.

Changes in the microtopography of the bone or tooth objects were tracked using quantitative 3DST[19] before and after extraction following established protocols[19,43]. Meshed axiomatic 3D models were generated and the following ISO 25178 parameters were used for statistical testing (paired Student's *t*-test, before and after the treatment, $\alpha \leq 0.01$): mean roughness (*Sa*), void volume (*Vvv*), peak curvature (*Spc*) and peak density (*Spd*). Further details regarding the 3DST measurements are provided in Supplementary Information 2.

In addition, we explored the compatibility of phosphate-based non-destructive DNA extraction with subsequent ^{14}C dating (Supplementary Information 3).

Non-destructive DNA isolation from artefacts

As sodium phosphate buffer did not cause substantial alterations of ancient bones and teeth in 3DST measurements, we modified a previously described method for temperature-controlled gradual DNA release from bone and tooth powder using this reagent[18] for non-destructive DNA extraction from complete bones and teeth. Stepwise extraction of DNA makes it possible to closely monitor the release of different DNA components during the extraction process (endogenous DNA, environmental DNA from the surrounding sediment, ancient human DNA and present-day contamination), potentially allowing inferences to be drawn about whether these components originate from traces of sediment that may still be adherent to the object, from its surface or its interior. We then applied this protocol to a total of 15 specimens from Quinçay, Bacho Kiro Cave and Denisova Cave. Temperature-controlled non-destructive DNA extraction was performed in an ancient DNA clean-room at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, according to the four steps below (see Fig. 1a for a schematic overview):

(1) Removal of sediment

This step was only performed for the freshly excavated material. First, clumps of sediment attached to the specimen were carefully removed by hand using a flexible disposable plastic microspatula. The specimen was then put into a 50-ml Falcon tube, rinsed by pouring between 20 ml and 50 ml of water into the tube, and transferred to a fresh Falcon tube. This procedure was repeated two to three times until no more sediment was released into the water. The tubes with water containing the sediment that had been washed off were then centrifuged at 16,400*g* for 5 min to pellet the sediment, and the clear supernatants were transferred to fresh tubes. All three types of material collected in this procedure (the manually removed sediment, the sediment pellet collected by centrifugation and the clear water) were subsequently subjected to DNA purification (see below).

(2) Temperature-controlled DNA release using sodium phosphate buffer

Each cleaned specimen was put into a 50-ml Falcon tube to which sodium phosphate buffer (0.5 M sodium phosphate, pH 7.0, and 0.1% Tween-20) was added until the specimen was completely submerged in the reagent (between 5 ml and 50 ml). After 30 min of incubation at room temperature (without agitation), the buffer was transferred to a fresh 50-ml Falcon tube. This step was repeated twice at room temperature (21 °C, for a total of three incubations) and then three times each at 37 °C, 60 °C and 90 °C (see above for details on the device and temperature settings used). A final incubation in water was performed at room temperature to remove residual reagent and the specimens were dried at room temperature for 5 days.

(3) DNA concentration

To facilitate subsequent DNA purification, half of the volumes of the water and phosphate buffer DNA fractions generated in steps (1) and (2) were reduced to between 50 µl and 75 µl by concentrating the DNA using Amicon Ultra-4 Centrifugal Filter Units with Ultracel-3 membranes (Millipore). For this, up to 4 ml or 15 ml of the respective sample was added to a filter unit, which was spun for 90 min at 4,000*g* in a centrifuge with an active cooling unit set to 21 °C. In cases in which the sample volume exceeded 4 ml or 15 ml, the flow-through was discarded and the filter unit was reloaded with remaining sample. Finally, a buffer exchange was performed by adding 4 ml

or 15 ml TE buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA) to the concentrated sample on top of the filter unit and spinning for 30 min at 4,000*g*. The supernatant was filled up to 300 µl with TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and then transferred into a fresh 1.5-ml Eppendorf low-bind tube and stored at −20 °C until further processing.

(4) DNA purification

DNA was isolated from the concentrated DNA fractions prepared in step (3) using a column-based method for silica-based ancient DNA extraction detailed elsewhere[44]. For this, 300 µl concentrated DNA was used as input for DNA purification using binding buffer 'D' and the purified DNA was recovered in 50 µl elution buffer. DNA extraction from the sediment samples removed from the artefacts manually or through water washes (sediment pellets) was performed using the same method but different input volumes. In brief, lysates were prepared by transferring up to 130 mg sediment to a 2-ml low-bind Eppendorf tube, adding up to 2 ml lysis buffer (0.45 M EDTA, pH 8.0, 0.05% Tween-20 and 0.25 mg ml^{-1} proteinase K) (1 ml for samples of less than 100 mg or 0.5 ml for samples of less than 25 mg) and incubating overnight at 37 °C under rotation. DNA purification was performed using 500 µl or 1,000 µl lysate.

Negative controls containing sodium phosphate buffer or lysis buffer without sample material were carried alongside the samples through all steps of subsequent sample preparation and sequencing. For a subset of samples, DNA was extracted from only the first phosphate buffer fraction produced at each of the four incubation temperatures. Supplementary Data 1 provides an overview of the DNA extracts generated in this study.

Library preparation

Of the extract, 10 µl was then converted into single-stranded DNA libraries using the automated protocol described in Gansauge et al. (2020)[23]. The number of unique library molecules obtained and the efficiency of library preparation were determined using two quantitative PCR assays[45]. Libraries were then amplified and double-indexed by PCR[46]. One library was prepared from each DNA fraction, except for the second and third 90 °C

phosphate fractions from the DCP1 from which five libraries were prepared, each to maximize the yield of sequence data. Negative controls containing no sample material were carried along each set of extraction and library preparation. Note that there was substantial variation in library preparation efficiency across samples (Fig. 2 and Supplementary Data 1), ranging from 21.8% on average for the phosphate DNA fractions obtained from Bacho Kiro Cave specimens to 60.8% for DCP1 and 62.0% for the Quinçay specimens.

Enrichment of mitochondrial and nuclear DNA by hybridization capture

To determine the taxonomic composition of the DNA recovered from the artefacts, libraries prepared from nearly all extracts generated in this study (all samples and controls, with the exception of the second and third phosphate fractions from BKP3) were enriched for mammalian mtDNA by hybridization capture with a probe set ('AA75') encompassing the mtDNA genomes of 242 mammalian species[24]. In cases in which several libraries were prepared from the same DNA extracts (that is, the 90 °C phosphate fractions from DCP1), only the first library was enriched for mammalian mtDNA. In addition, all libraries were enriched specifically for hominin mtDNA using a probe set ('AA163') designed in 1-bp tiling based on the revised Cambridge reference sequence of the human mitochondrial genome[32]. Both types of mtDNA captures were performed using two consecutive rounds of automated capture as detailed in Slon et al. (2017)[6] or Zavala et al. (2022)[47]. An overview of the capture reactions performed is provided in Supplementary Data 1.

In addition to mtDNA captures, 11 libraries prepared from the first, second and third 90 °C phosphate fractions of DCP1 were enriched for human nuclear DNA by two consecutive rounds of in-solution capture[48]. This enrichment was performed using a subset of a previously designed capture probe panel[8], named AA204, targeting two groups of SNPs: (1) 59,232 SNPs, randomly selected from set 6 'hominin diagnostic sites'[8], which represent positions in the genome where primates differ from other mammals and are used to quantify faunal mis-alignments; and (2) 411,492 SNPs selected from the '1240k' panel[30,49] (set 2 in Vernot et al. (2021)[8]).

This set of SNPs is informative for investigating modern human population histories. Both SNP groups are located in regions of large evolutionary sequence divergence between humans and other mammals[8].

Sequencing and raw sequence processing

Libraries enriched for mtDNA were combined into pools and sequenced on multiple lanes of a MiSeq sequencer (Illumina Technologies) in paired-end configuration with two index reads (2× 76 + 2× 8 cycles). Libraries enriched for human nuclear DNA were sequenced either on a HiSeq 2500 using the same configuration or on a HiSeq4000 (both Illumina Technologies) in single-read configuration with two index reads (1× 76 + 2× 8 cycles). In addition, one library prepared from the second 90 °C phosphate fraction of the DCP1 was sequenced directly (shotgun sequencing, without hybridization capture) on one lane of a HiSeq 4000 sequencer (Illumina Technologies) in single-read configuration. Base calling was performed using Illumina's Bustard tool and sequences were assigned to the library that they originated from, requiring perfect matches to the expected index combinations. LeeHom (https://github.com/mpieva/leeHom/tree/v.1.1.5)[50] was used to trim adapters and, for paired-end data, to merge overlapping paired-end reads.

Taxonomic assignment of mtDNA sequences

Sequences resulting from mammalian or human mtDNA capture were assigned to mammalian taxa on the biology family level using a previously published computational pipeline[6] based on BLAST and MEGAN (version 0.0.12)[51], with modifications in data filtering detailed in Vernot et al. (2021)[8] (Supplementary Data 1 and Supplementary Information 4). In brief, as described in the latter study, false identifications of taxa were minimized by requiring at least three unique sequences (covering at least 105 positions in the reference genome) to be assigned to a family, and by requiring these sequences to represent at least 1% of all taxonomically identified sequences. The presence of ancient DNA was determined individually for the sequences from each family (and each library) by computing the frequency of terminal C-to-T substitutions as a proxy for the deamination rate at the molecule ends. Substitution frequencies significantly higher than 10% on both

molecule ends (based on 95% binomial CIs) were then taken as evidence for the presence of ancient DNA from the respective family.

Human and cervid mtDNA analysis

Present-day human contamination in the fractions from DCP1 that yielded ancient human mtDNA was estimated using the software tool AuthentiCT (version 1.0.0)[52] and a near-complete consensus sequence called for the second 90 °C fraction from DCP1 using positions with at least tenfold coverage of the mtDNA genome. This consensus sequence was then used for haplogroup assignment with Haplogrep 2 (version2.4.0)[53] and to identify 'diagnostic' positions in the mtDNA genome, which were used to determine the support for the consensus sequence in each of the DNA fractions recovered from DCP1 (Extended Data Table 3). Tree building and genetic dating were performed using BEAST2 (version 2.6.6)[54]. Further information on the mtDNA analyses are provided in Supplementary Information 5. For genetic dating of the deer DNA component, near-complete wapiti mtDNA genomes were reconstructed from the first 60 °C phosphate fraction obtained from DCP1, as well as eight ancient wapiti samples as described in detail in Supplementary Information 6.

Human nuclear DNA analysis

Human DNA capture data were processed as previously described[8] and data from the libraries with the lowest estimates of modern human contamination (based on AuthentiCT) were merged for further analyses. Principal component analysis including sequence data from present-day and other ancient human individuals was performed using smartpca (from EIGENSOFT package version 8.0.0)[55]. Using the R package admixr (version 0.7.1)[28], $f3$-statistics were calculated to determine shared genetic drift between DCP1 and a selection of modern and ancient human populations and their relationships were further evaluated using D-statistics. Sex determination for the human DNA component recovered from DCP1 was performed by comparing the coverage of the X chromosome and the autosomes in shotgun data obtained from the second 90 °C phosphate fraction after filtering against faunal mis-mappings[8]. Further details are provided in Supplementary Information 7.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are included in the Article and its Supplementary Information.

References

1. Bailey, G. Time perspectives, palimpsests and the archaeology of time. *J. Anthropol. Archaeol.* **26**, 198–223 (2007).

2. Kuhn, S. & Stiner, M. What's a mother to do? The division of labor among Neandertals and modern humans in Eurasia. *Curr. Anthropol.* **47**, 953–981 (2006).

3. Nowell, A. *Growing up in the Ice Age: Fossil and Archaeological Evidence of the Lived Lives of Plio-Pleistocene Children* (Oxbow Books, 2021).

4. Owen, L. R. *Distorting the Past: Gender and the Division of Labor in the European Upper Paleolithic* (Kerns Verlag, 2005).

5. Villotte, S., Churchill, S. E., Dutour, O. J. & Henry-Gambier, D. Subsistence activities and the sexual division of labor in the European Upper Paleolithic and Mesolithic: evidence from upper limb enthesopathies. *J. Hum. Evol.* **59**, 35–43 (2010).

6. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017).

7. Zavala, E. I. et al. Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature* **595**, 399–403 (2021).

8. Vernot, B. et al. Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science* **372**, eabf1667 (2021).

9. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).

10. Brundin, M., Figdor, D., Sundqvist, G. & Sjogren, U. DNA binding to hydroxyapatite: a potential mechanism for preservation of microbial DNA. *J. Endod.* **39**, 211–216 (2013).

11. Korlevic, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).

12. Rohland, N., Siedel, H. & Hofreiter, M. Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *Biotechniques* **36**, 814–821 (2004).

13. Gomes, C. et al. Nondestructive extraction DNA method from bones or teeth, true or false? *Forensic Sci. Int. Genet.* **5**, e279–e282 (2015).

14. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007).

15. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).

16. Gamba, C. et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* **16**, 459–469 (2016).

17. Salamon, M., Tuross, N., Arensburg, B. & Weiner, S. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc. Natl Acad. Sci. USA* **102**, 13783–13788 (2005).

18. Essel, E., Korlevic, P. & Meyer, M. A method for the temperature-controlled extraction of DNA from ancient bones. *Biotechniques* **71**,

382–386 (2021).

19. Schulz, E., Calandra, I. & Kaiser, T. M. Applying tribology to teeth of hoofed mammals. *Scanning* **32**, 162–182 (2010).

20. Schulz-Kornas, E. et al. Everything matters: molar microwear texture in goats (*Capra aegagrus hircus*) fed diets of different abrasiveness. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **552**, 109783 (2020).

21. Roussel, M., Soressi, M. & Hublin, J. J. The Chatelperronian conundrum: blade and bladelet lithic technologies from Quincay, France. *J. Hum. Evol.* **95**, 13–32 (2016).

22. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).

23. Gansauge, M. T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).

24. Slon, V. et al. Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quatern. Int.* **398**, 210–218 (2015).

25. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).

26. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).

27. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

28. Petr, M., Vernot, B. & Kelso, J. admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).

29. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).

30. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).

31. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

32. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).

33. Green, R. E. et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).

34. Hajdinjak, M. et al. Reconstructing the genetic history of late Neanderthals. *Nature* **555**, 652–656 (2018).

35. Kay, P. C.A. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).

36. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

37. Welker, F. et al. Variations in glutamine deamidation for a Chatelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. *Sci. Technol. Archaeol.* **3**, 15–27 (2017).

38. Rendu, W. et al. Subsistence strategy changes during the Middle to Upper Paleolithic transition reveals specific adaptations of human populations to their environment. *Sci. Rep.* **9**, 15817 (2019).

39. Fewlass, H. et al. A ^{14}C chronology for the Middle to Upper Palaeolithic transition at Bacho Kiro Cave, Bulgaria. *Nat. Ecol. Evol.* **4**, 794–801 (2020).

40. Hublin, J.-J. et al. Initial Upper Palaeolithic *Homo sapiens* from Bacho Kiro Cave, Bulgaria. *Nature* **581**, 299–302 (2020).

41. Tsanova, T. et al. The Initial Upper Palaeolithic lithic assemblage from Bacho Kiro Cave (Bulgaria). In *9th Annual Meeting of the European Society for the Study of Human Evolution*. 120 (2020).

42. Hofreiter, M. Nondestructive DNA extraction from museum specimens. *Methods Mol. Biol.* **840**, 93–100 (2012).

43. Martisius, N. L. et al. Time wears on: assessing how bone wears using 3D surface texture analysis. *PLoS ONE* **13**, e0206078 (2018).

44. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).

45. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* **27**, 1230–1237 (2017).

46. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).

47. Zavala, E. I. et al. Quantifying and reducing cross-contamination in single- and multiplex hybridization capture of ancient DNA. *Mol. Ecol. Resour.* **22**, 2196–2207 (2022).

48. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).

49. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).

50. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).

51. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).

52. Peyregne, S. & Peter, B. M. AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biol.* **21**, 246 (2020).

53. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).

54. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

55. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

56. R Core Team. *R: A Language and Environment for Statistical Computing*. http://www.R-project.org/ (R Foundation for Statistical Computing, 2013).

Acknowledgements

We thank V. Slon, C. Hopfe, M. Peresani and W. Roebroeks for their contributions in an early exploratory phase of the project; P. Kosintsev for providing samples; A. Hübner for help with data analysis; B. Schellbach and A. Weihmann for help with sequencing; S. Nagel, J. Richter, B. Nickel and A. Aximu-Petri for help in the laboratory; J. Visagie for data processing; L. Jauregui, A. Bossoms Mesa, A. Pelin Sümer, H. Temming, A. Lister, M. Meiri and Z. Rezek for comments on the manuscript and/or logistical support; the National Museum of Natural History in Sofia for cooperating; R. Spasov for help with excavation at Bacho Kiro Cave; F. Lévêque, A. Lévêque, C.-H. Bachelier and J. Bachelier for facilitating and supporting research at Quinçay and Les Cottés; S. Rigaud for help with morphological identification of Quinçay samples; the Max Planck Society for funding; and the French Ministry of Culture for allowing and funding research at Les

Cottés and Quinçay. The archaeological studies at Denisova Cave were funded by the Russian Science Foundation (no. 22-28-00049). M.Soressi is funded by the NWO VICI award (Neandertal Legacy VI.C.191.07). The radiocarbon dating work received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 715069 (FINDER) to K.D. Work by E.I.Z. was partially funded by the Miller Institute for Basic Research in Science, University of California Berkeley.

Funding

Open access funding provided by Max Planck Society.

Author information

Author notes

1. Elena I. Zavala

   Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

2. Tsenka Tsanova

   Present address: Department of Chemistry "Giacomo Ciamician", Alma Mater Studiorum, University of Bologna, Bologna, Italy

3. Jean-Jacques Hublin

   Present address: Chaire de Paléoanthropologie, Collège de France, Paris, France

4. Mateja Hajdinjak

   Present address: Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

5. These authors contributed equally: Elena Essel, Elena I. Zavala, Ellen Schulz-Kornas, Marie Soressi, Matthias Meyer

Authors and Affiliations

1. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

   Elena Essel, Elena I. Zavala, Helen Fewlass, Benjamin Vernot, Anna Schmidt, Merlin Szymanski, Tsenka Tsanova, Shannon P. McPherron, Jean-Jacques Hublin, Janet Kelso, Svante Pääbo & Matthias Meyer

2. Department of Biology, San Francisco State University, San Francisco, CA, USA

   Elena I. Zavala

3. Department of Cariology, Endodontology and Periodontology, University of Leipzig, Leipzig, Germany

   Ellen Schulz-Kornas

4. Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia

   Maxim B. Kozlikin, Michael V. Shunkov & Anatoly P. Derevianko

5. Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Vienna, Austria

   Katerina Douka

6. Human Evolution and Archaeological Sciences (HEAS) Research Network, University of Vienna, Vienna, Austria

   Katerina Douka

7. Earth Sciences Department, Natural History Museum, London, UK

Ian Barnes

8. Maison de la Recherche, Université de Toulouse-Jean Jaurès, CNRS UMR 5608 TRACES, Toulouse, France

Marie-Cécile Soulier

9. National Institute of Archaeology with Museum, Bulgarian Academy of Sciences, Sofia, Bulgaria

Nikolay Sirakov

10. National Museum of History, Sofia, Bulgaria

Elena Endarova

11. Ancient Genomics Laboratory, The Francis Crick Institute, London, UK

Mateja Hajdinjak

12. Faculty of Archaeology, Leiden University, Leiden, The Netherlands

Marie Soressi

Contributions

E.Essel, M.Soressi and M.M. designed the study. E.Essel, E.S.-K., H.F., M.Soressi and M.M. developed and tested methods. M.B.K., M.V.S., A.P.D., I.B., M.-C.S., T.T., N.S., E.Endarova, S.P.M., J.-J.H. and M.Soressi excavated, identified or characterized archaeological material and/or provided archaeological context and interpretation. E.Essel, E.S.-K., H.F., K.D. and A.S. performed laboratory experiments or collected data. E.Essel, E.I.Z., E.S.-K., H.F., B.V., K.D., M.Szymanski, J.K., S.P., M.H., M.Soressi and M.M. performed, aided in or supervised data analysis. E.Essel, E.I.Z., M.Soressi and M.M. wrote the manuscript with input from all other authors. E.Essel, E.I.Z., E.S.-K., M.M. and M.Soressi contributed equally to this work.

Corresponding authors

Correspondence to Elena Essel, Marie Soressi or Matthias Meyer.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Sophie Archambault de Beaune, Carles Lalueza-Fox and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Microtopographic surface measurements.

Boxplots of the four selected surface texture parameters *Sa* (arithmetical mean height), *Vvv* (void volume of the valleys), *Spc* (arithmetic mean peak curvature), and *Spd* (density of peaks) by treatment (EDTA = ethylenediaminetetraacetic acid, GuSCN = guanidine thiocyanate reagent, Phos. = sodium phosphate buffer supplemented with detergent, Bleach = sodium hypochlorite, water = distilled water; e = enamel, d = dentine).

Extended Data Fig. 2 Faunal remains used in reagent testing.

Photographs of samples taken before and after treatment with various reagents used in ancient DNA extraction (GuSCN = guanidine thiocyanate reagent, EDTA = ethylenediaminetetraacetate solution, Phosphate = sodium phosphate buffer with detergent, Bleach = sodium hypochlorite solution). The black bar represents 1 cm. Note that colors are not directly comparable, as the photographs were taken at slightly different angles, with different light settings and camera adjustments.

Extended Data Fig. 3 Artefacts before and after DNA extraction.

Photographs of samples taken before and after phosphate-based, non-destructive DNA extraction. The black bar represents 1 cm. Note that colors are not directly comparable, as the photographs were taken at slightly different angles, with different light settings and camera adjustments.

Extended Data Fig. 4 Photograph of DCP1 as it became exposed during excavation.

The photograph was taken shortly before the pendant was removed and placed into a plastic bag using gloves.

Extended Data Fig. 5 Shared genetic drift between DCP1 and different present-day and ancient human individuals.

Each point represents a distinct statistic using the calculation f3(X,Y; Mbuti) where Mbuti serves as an outgroup, population X is an ancient or present-day human population and Y is DCP1. Warmer colors on the map[56] represent more shared genetic drift.

Extended Data Fig. 6 Genetic affinity of DCP1 to the genomes of other modern and ancient humans.

(A) Genetic affinities of DCP1 and Ancient North Eurasian (ANE) individuals (Y) to different ancient Siberians (W and X) based on 9,297-787,534 SNPs from the “1240k” SNP panel. **(B)** D-statistics comparing the

genetic affinity of DCP1 and Mal'ta1 to a selection of ancient modern humans based on 13,618-237,630 SNPs from the "1240k" SNP panel. Red color indicates all sequences, grey color deaminated sequences only. **(C)** Genetic affinities of DCP1 and ANE individuals (Y) to different ancient Americans (W and X) based on 42,548-574,391 SNPs from the "1240k" SNP panel. The calculations were performed using ADMIXTOOLS (version 5.1) via admixr (version 0.7.1) using all fragments or deaminated fragments only for DCP1. The error bars represent one standard error as calculated by a Weighted Block Jackknife. AG3 = Afontova Gora 3.

Extended Data Table 1 Description of the samples used in reagent testing and bone/tooth artefacts used for non-destructive DNA extraction
Extended Data Table 2 Differences between the mtDNA consensus sequence of DCP1 and other human mtDNA genomes
Extended Data Table 3 Support for the mtDNA consensus sequence reconstructed from DCP1

Supplementary information

Supplementary Information

Reporting Summary

Peer Review File

Supplementary Data file 1

Overview of DNA lysates, extracts and libraries prepared in this study and results of the mammalian and human mtDNA captures.

Supplementary Data file 2

Overview and summary statistics of the human nuclear DNA captures targeting 470,724 positions in the genome.

Supplementary Data file 3

Calculated D-statistics by ADMIXTOOLS via admixr using D(W, X; ANE, Chimp) where W is a present-day African and X is a present-day non-African. The values in this table are visualized in Supplementary Figure 7.5.

Supplementary Data file 4

Calculated D-statistics by ADMIXTOOLS via admixr using D(W, X; ANE, Mbuti) where W is a present-day Americans and X is a present-day non-African/American. The values in this table are visualized in Supplementary Figure 7.6.

Supplementary Data file 5

Calculated D-statistics by ADMIXTOOLS via admixr using D(W, X; ANE, Mbuti) where W is an ancient Siberian and X is an ancient American. The values in this table are visualized in Supplementary Figure 7.7.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Essel, E., Zavala, E.I., Schulz-Kornas, E. *et al.* Ancient human DNA recovered from a Palaeolithic pendant. *Nature* **618**, 328–332 (2023). https://doi.org/10.1038/s41586-023-06035-2

- Received: 28 November 2022
- Accepted: 30 March 2023
- Published: 03 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06035-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Human DNA from a prehistoric artefact**
  - Michael Attwaters

  Nature Reviews Genetics (2023)

Ancient woman’s DNA recovered from a 20,000-year-old pendant

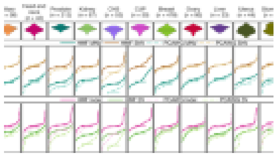
Nature Research Briefing 03 May 2023

Prehistoric pendant’s DNA reveals the person who held it

- Elissa Welle

Nature News 03 May 2023

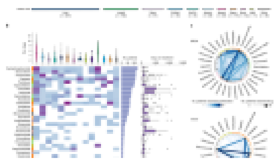
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06035-2



Pan-cancer whole-genome analyses of metastatic solid tumours

23 October 2019

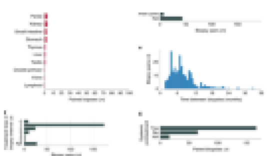
Peter Priestley, Jonathan Baber, … Edwin Cuppen



Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes

13 April 2020

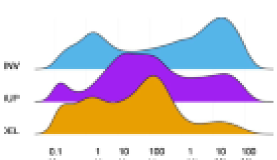
Erin Pleasance, Emma Titmuss, … Marco A. Marra



Limited evolution of the actionable metastatic cancer genome under therapeutic pressure

09 August 2021

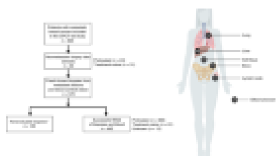
Joris van de Haar, Louisa R. Hoes, … Emile E. Voest



Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features

25 January 2021

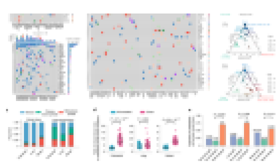
Pauline A. J. Mendelaar, Marcel Smid, … Saskia M. Wilting



The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies

30 September 2019

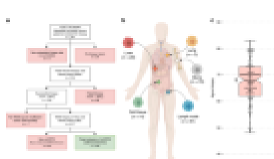
Lindsay Angus, Marcel Smid, … John W. M. Martens



Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases

18 May 2020

Zheng Hu, Zan Li, … Christina Curtis



The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact

20 November 2019

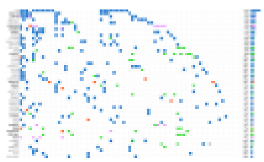
Lisanne F. van Dessel, Job van Riet, … Martijn P. Lolkema



[Automated next-generation profiling of genomic alterations in human cancers]

20 May 2022

Laurel A. Keefer, James R. White, … Mark Sausen



[Comprehensive genomic and epigenomic analysis in cancer of unknown primary guides molecularly-informed therapies despite heterogeneity]

02 August 2022

Lino Möhrmann, Maximilian Werner, … Hanno Glimm

- Article
- Open Access
- Published: 10 May 2023

Pan-cancer whole-genome comparison of primary and

metastatic solid tumours

- Francisco Martínez-Jiménez,
- Ali Movasati,
- Sascha Remy Brunner,
- Luan Nguyen,
- Peter Priestley,
- Edwin Cuppen &
- …
- Arne Van Hoeck

Show authors

Nature **volume 618**, pages 333–341 (2023)

- 29k Accesses
- 1 Citations
- 367 Altmetric
- Metrics details

Abstract

Metastatic cancer remains an almost inevitably lethal disease[1,2,3]. A better understanding of disease progression and response to therapies therefore remains of utmost importance. Here we characterize the genomic differences between early-stage untreated primary tumours and late-stage treated metastatic tumours using a harmonized pan-cancer analysis (or reanalysis) of two unpaired primary[4] and metastatic[5] cohorts of 7,108 whole-genome-sequenced tumours. Metastatic tumours in general have a lower intratumour heterogeneity and a conserved karyotype, displaying only a modest increase in mutations, although frequencies of structural variants are elevated overall. Furthermore, highly variable tumour-specific contributions of mutational footprints of endogenous (for example, SBS1 and APOBEC) and exogenous

mutational processes (for example, platinum treatment) are present. The majority of cancer types had either moderate genomic differences (for example, lung adenocarcinoma) or highly consistent genomic portraits (for example, ovarian serous carcinoma) when comparing early-stage and late-stage disease. Breast, prostate, thyroid and kidney renal clear cell carcinomas and pancreatic neuroendocrine tumours are clear exceptions to the rule, displaying an extensive transformation of their genomic landscape in advanced stages. Exposure to treatment further scars the tumour genome and introduces an evolutionary bottleneck that selects for known therapy-resistant drivers in approximately half of treated patients. Our data showcase the potential of pan-cancer whole-genome analysis to identify distinctive features of late-stage tumours and provide a valuable resource to further investigate the biological basis of cancer and resistance to therapies.

Main

Metastatic spread involves tumour cell detachment from a primary tumour, colonization of a secondary tissue and growth in a hostile environment[1,2]. Advanced metastatic tumours are frequently able to resist aggressive treatment regimes[6]. Despite the many efforts to understand these phenomena[3,7,8,9], we still have limited knowledge of the contribution of genomic changes that equip tumours with these extraordinary capacities. Thus, it is essential to characterize genomic differences between primary and metastatic cancers and quantify their effect on therapy resistance to understand and harness therapeutic interventions that establish more effective and more personalized therapies[10].

Although extensive whole-genome analyses of primary or metastatic tumour types have been conducted[4,5], large-scale comparative studies between the two tumour stages remain limited due to the logistical challenges associated with obtaining pan-cancer cohorts of primary and metastatic cancers. To circumvent this issue, most comparison studies have relied on unpaired whole-exome data or have adopted more targeted approaches with a specific focus on driver gene landscapes[11,12,13]. However, these efforts have frequently involved separated processing pipelines for primary and metastatic cohorts, complicating the analysis of genomic features that are

highly sensitive to the selected data-processing strategy[14,15]. A recent study that uniformly analysed more than 25,000 tumours[16] has provided a comprehensive overview of the genomic differences, driver alteration patterns and organotropism using clinical gene-panel sequencing as a base. However, this genomic analysis approach prevented the exploration of the full spectrum of genomic alterations that have a role in tumorigenesis, such as structural variation and mutational scarring.

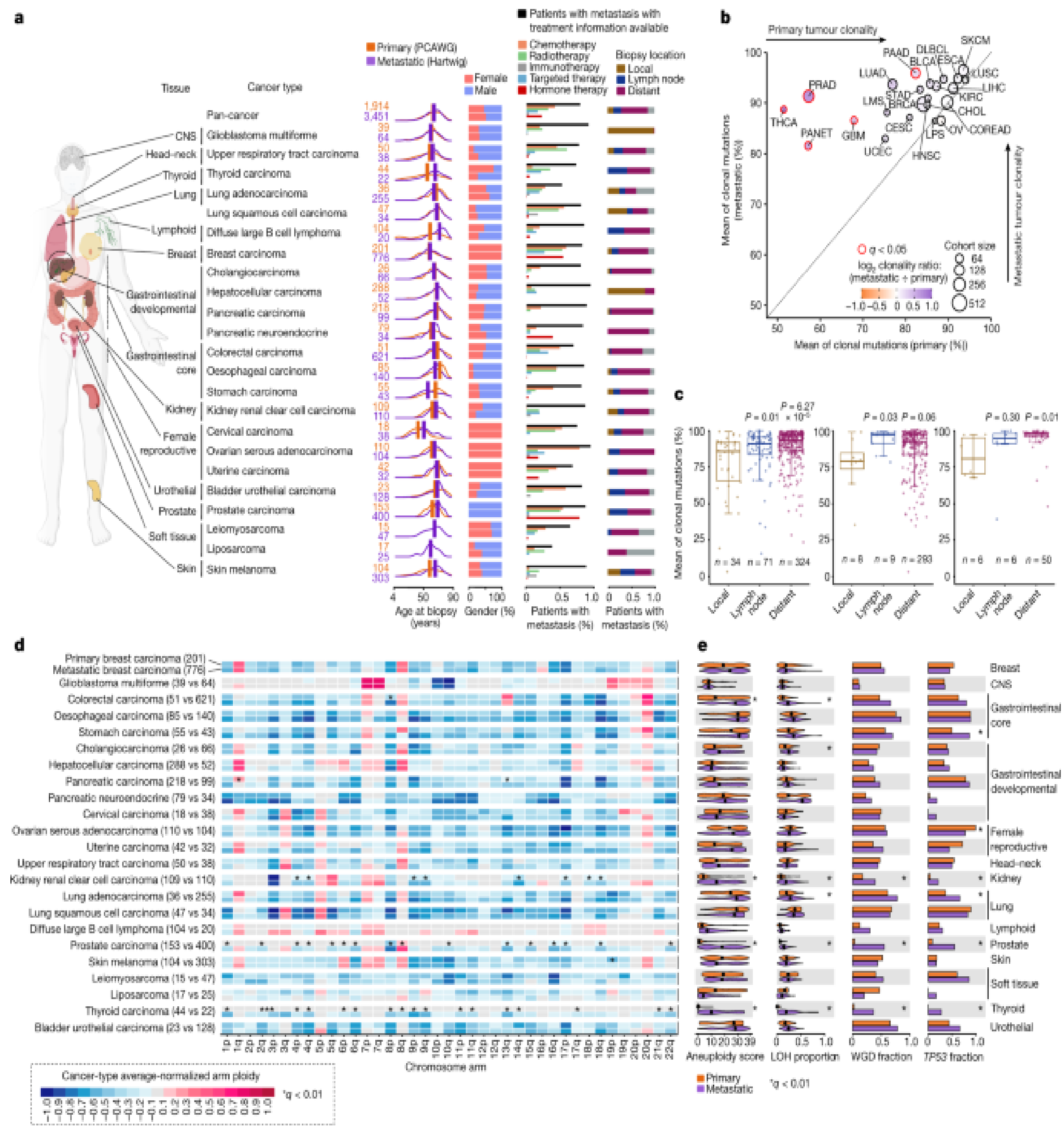
Harmonized whole-genome-sequenced tumours

Here we created a uniformly processed whole-genome-sequenced (WGS) inventory of 7,108 matched tumour and normal genomes from two unpaired primary and metastatic cohorts. We first collated the Hartwig Medical Foundation (Hartwig) dataset[5], which included 4,784 samples from 4,375 patients with metastatic cancers. Then, we reprocessed 2,835 primary tumour samples from the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium[4] using the open-source Hartwig analytical pipeline[5,17] to harmonize somatic calling and annotations of events, and to eliminate processing biases (Extended Data Fig. 1 and Supplementary Table 1). Reassuringly, per-sample comparison of the number of single-base substitutions (SBSs), double-base substitutions (DBSs), indels (IDs) and structural variants (SVs) revealed a strong agreement between our results and the consensus calls originally generated by the PCAWG consortium (Supplementary Note 1). In addition, our processing pipeline strategy was minimally affected by differences in sequencing coverage, enabling a reasonable comparison of WGS samples from heterogeneous sources (Supplementary Note 1). A total of 7,108 tumour samples from 71 cancer types met the processing pipeline quality standards (Methods and Extended Data Fig. 1a) and constitutes one of the largest publicly available datasets of WGS tumours.

We next focused on 23 cancer types from 14 tissues with sufficient sample representation, comprising 5,365 tumour samples (1,914 primary and 3,451 metastatic) to explore genomic differences between primary and metastatic tumours (Fig. 1a and Extended Data Fig. 1a). Within this dataset, patients with metastatic tumours were slightly older at biopsy than patients with primary tumours (mean of 1.67 years older across all cancer types), although

patients with metastatic prostate and thyroid carcinomas, and diffuse large B cell lymphoma were markedly older than their primary counterparts. Consistent gender proportions were observed across all cancer types except for thyroid adenocarcinomas (metastatic: 72% male and 28% female; primary: 25% male and 75% female). Treatment information was available for 83.7% of patients with metastatic tumours, which is essential to gauge specific treatment-induced contributions to genomic differences between primary and metastatic tumours (Fig. 1a). Finally, biopsy locations were annotated for 84.2% of patients with metastatic tumours (12.2% from metastatic lesions in the primary tissue (local), 16.2% in lymph nodes and 55.7% in distant locations) and displayed a highly tumour-type-specific distribution pattern, probably reflecting both the dissemination patterns of the tumours and the accessibility for safe clinical sampling.

Fig. 1: Database overview and global genomic features.



a, Anatomical location of the 23 cancer types, ordered by tissue or origin, included in this study. From left to right: sample size, age at biopsy, gender, treatment type and biopsy site of the cohort with metastatic tumours. CNS, central nervous system. The image in **a** was created using BioRender (https://biorender.com). **b**, Mean percentage of clonal mutations in primary (*x* axis) and metastatic (*y* axis) tumours. The dots are coloured according to the $\log_2$ of the clonality ratio (metastatic divided by primary). The size of the dots is proportional to the total number of samples (primary and metastatic). The red edge lines represent a two-sided Mann–Whitney adjusted $P < 0.05$.

BLCA, bladder urothelial carcinoma; BRCA, breast carcinoma; CESC, cervical carcinoma; CHOL, cholangiocarcinoma; COREAD, colorectal carcinoma; DLBCL, diffuse large B cell lymphoma; ESCA, oesophageal carcinoma; GBM, glioblastoma multiforme; HNSC, upper respiratory tract carcinoma; KIRC, kidney renal clear cell carcinoma; LIHC, hepatocellular carcinoma; LMS, leiomyosarcoma; LPS, liposarcoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; PAAD, pancreatic carcinoma; PANET, pancreatic neuroendocrine tumour; PRAD, prostate carcinoma; OV, ovarian serous adenocarcinoma; SKCM, skin melanoma; STAD, stomach carcinoma; THCA, thyroid carcinoma; UCEC, uterine carcinoma. **c**, Tumour clonality according to the metastatic biopsy location in breast (left), colorectal (middle) and oesophageal (right) carcinomas. *n* Refers to the number of samples. *P* refers to Mann–Whitney two-sided *P* value. For boxplots, the centre line indicates the median; the box limits denote the first and third quartiles; and the whiskers indicate the lowest or highest data points at the first quartile minus or plus 1.5× the interquartile range. **d**, Heatmap representing the normalized mean chromosome arm ploidy gains and losses relative to the expected $2n$ ploidy status in primary (top) and metastatic (bottom) tumours. *Adjusted $P < 0.01$ (two-sided Mann–Whitney). **e**, Comparison of four genomic instability indicators between primary (top) and metastatic (bottom) tumours. From left to right: aneuploidy score from ref. [20], the proportion of genome undergoing LOH, and the fraction of samples bearing whole-genome duplication (WGD) and *TP53* alterations. The black dots represent the median values. *Adjusted $P <$ 0.01 using two-sided Fisher's exact test for WGD and *TP53*, and two-sided Mann–Whitney test for the continuous features.

Comparison of global genomic features

We first explored global genomic differences between primary and metastatic tumours across the aforementioned 23 cancer types. Metastatic tumours showed an overall increase in clonality compared with their primary tumour counterparts (Fig. 1b). Particularly, five cancer types had a significantly higher metastatic average clonality ratio, ranging from 13.6% increased mean clonality in pancreatic carcinoma to 37.2% in thyroid carcinoma. Within the group of patients with metastatic breast carcinoma, distant and lymph node tumour biopsies showed significantly higher

clonality ratios than local metastatic lesions (Fig. 1c and Supplementary Note 2). This increase in clonality was also observed in distant tumour biopsies of oesophageal and colorectal carcinomas (Fig. 1c). Nevertheless, the biopsy location did not influence tumour clonality in other cancer types such as lung adenocarcinoma and skin melanoma (Extended Data Fig. 1b), suggesting that patterns of tumour dissemination are highly tumour-type specific[9]. Our results support the model that metastatic lesions generally have lower intratumour heterogeneity[16], which may be explained by a single major subclone seeding event from the primary cancer and/or by severe evolutionary constraints imposed by anticancer therapies.

Comparison of chromosome arm aneuploidy profiles revealed a generally conserved portrait, which was strongly shaped by the cell of origin (Fig. 1d and Supplementary Table 2), supporting the notion that tumour karyotype is generally defined at early stages of tumorigenesis[18]. Only metastatic kidney renal clear cell, prostate and thyroid carcinomas showed substantial changes compared with primary tumours, encompassing 91% (43 of 47) of all significant discrepancies. Besides the poor prognostic marker 8q gain in metastatic prostate carcinoma[19], all discrepancies were associated with an increased prevalence of chromosomal arm losses at the metastatic setting. Remarkably, 30% (14 of 46) of the metastatic-enriched chromosome arm losses were retained when comparing non-WGD tumours (Extended Data Fig. 1c,d), indicating that other factors, aside from whole-genome doubling[20,21], have an important role in the accumulation of arm-sized chromosomal aberrations.

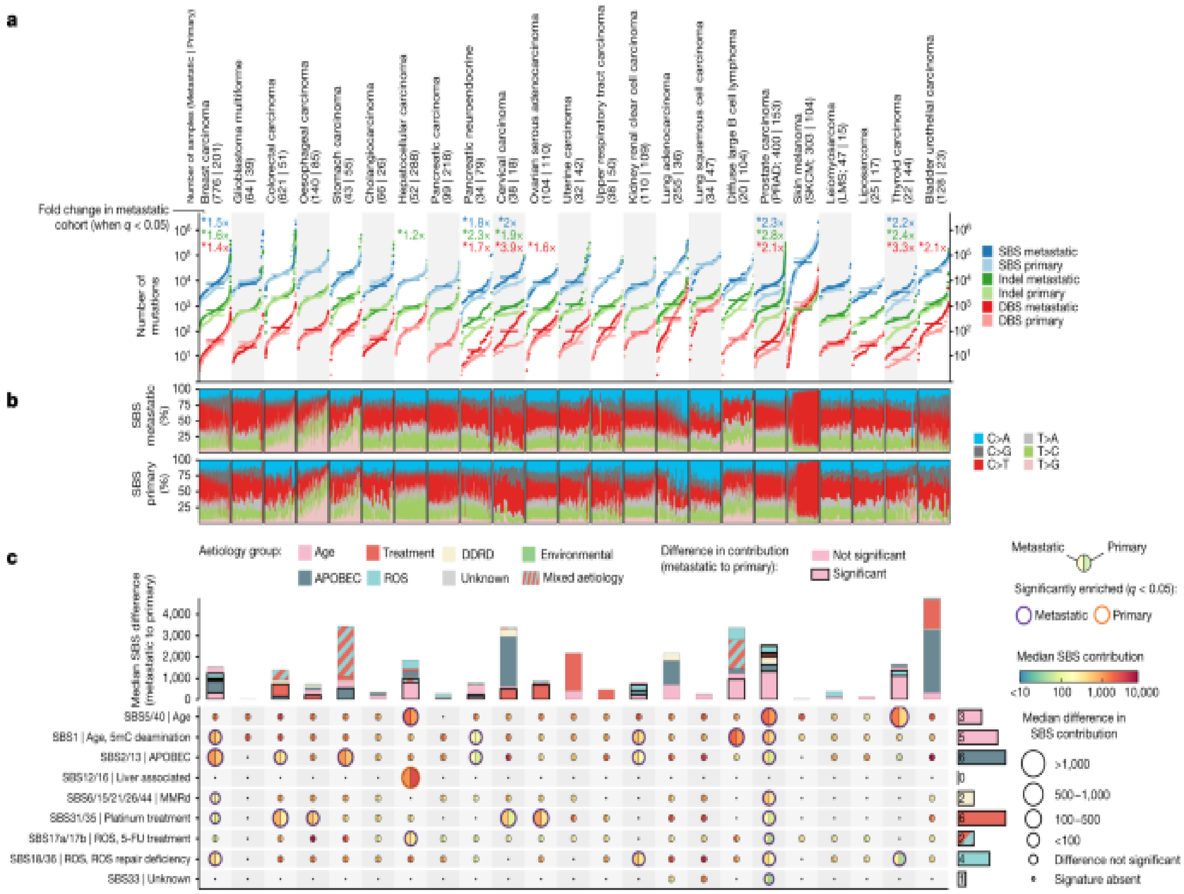
The same three cancer types also showed persistent increases in specific genomic instability indicators (that is, chromosomal aneuploidy score[20], loss of heterozygosity (LOH) genome fraction, WGD[22] and *TP53* alterations[23]) in the metastatic cohort (Fig. 1e and Supplementary Table 2). Although all of these indicators are elevated pan-cancer in WGD tumours (Extended Data Fig. 1c,d), non-WGD metastatic tumours of these three cancer types also had significantly greater aneuploidy and LOH scores. Furthermore, patients with metastatic tumours from other cancer types, including lung and colorectal adenocarcinomas, also displayed a moderate increase in aneuploidy and LOH scores, although they seemed to be primarily associated with higher metastatic WGD rates (Extended Data Fig. 1c). Our results thus revealed

that the majority of cancer types have already acquired variable degrees of chromosomal arm aneuploidy in early stages of tumorigenesis. However, in certain cancers, such as kidney renal clear cell, prostate and thyroid carcinomas, significantly increased levels of genomic instability were induced in later evolutionary stages, which were, in turn, associated with substantial additional karyotypic changes.

Tumour mutation burden

We observed that the small variant tumour mutation burden (TMB), collectively encompassing SBSs, DBSs and IDs, was only moderately increased in metastatic tumours compared with primary tumours across the 23 cancer types tested (fold-change increases of 1.25 ± 0.47 for SBSs, 1.55 ± 0.86 for DBSs and 1.45 ± 0.53 for IDs; mean ± standard deviation (s.d.)). In fact, 15 of the 23 cancer types had no significant increase in mutation burden for any mutation type. Only five cancer types (breast, cervical, thyroid and prostate carcinomas and pancreatic neuroendocrine tumour) had a consistent increase for the three mutation types at the metastatic stage, although the mutation profiles lacked systematic differences between primary and metastatic tumours (Fig. 2b and Extended Data Fig. 2a). Finally, further TMB comparisons grouping by tumour subtypes, metastatic biopsy locations and primary clinical progression status generally provided consistent results, although cancer-type-specific particularities are present (Supplementary Note 2). These results show that TMB is not necessarily indicative of tumour progression status and that the overall mutational spectra are tightly shaped by the mutational processes that were already active before and during primary tumour development.

Fig. 2: TMB and mutational processes.



a, Cumulative distribution function plot (samples were ranked independently for each variant type) of TMB for each cancer type for SBS (blue), IDs (green) and DBS (red). The horizontal lines represent median values. The fold-change labels are included only when two-sided Mann–Whitney comparison renders a significant adjusted $P < 0.05$. **b**, SBS mutational spectra of patients with metastatic (top) and primary (bottom) tumours. Patients are ordered according to their TMB. DDRD, DNA damage repair deficiency; ROS, reactive oxygen species. **c**, Moon plot representing the SBS mutational burden differences attributed to each mutational signature in metastatic (main plot, left) and primary (main plot, right) tumours. The edge thickness and colours represent significant differences (two-sided Mann–Whitney adjusted $P < 0.05$, ±1.4× fold change) and the direction of the enrichment, respectively. The size of the circles are proportionate to the mutation burden difference. The bars on the right indicate the number of metastatic cancer types with a mutational signature with significant enrichment. The top stacked bars represent the cumulative signature exposure difference. The thicker bar edge lines represent significance. Bars

are coloured according to the annotated aetiology. Only mutational signatures with known aetiology or with at least one cancer type with significant metastatic enrichment are included. 5-FU, 5-fluorouracil; 5mC, 5-methylcytosine; MMRd, mismatch repair deficiency.

Mutational processes activity comparison

To determine whether the TMB differences may be attributed to differential activity of environmental or endogenous mutational processes, we assessed the activities of all operative mutational signatures in a quantitative and relative manner. We found that mutations attributed to cytotoxic treatments were significantly enriched in ten cancer types (Fig. 2c (red top bars) and Extended Data Figs. 2b,c and 3a for relative contributions). Platinum-based chemotherapies (SBS31/SBS35 and DBS5) showed the strongest mutagenic effect with 551 ± 575 (mean ± s.d.) SBS mutations and 32 ± 22 (mean ± s.d.) DBS-attributed mutations on average per sample. In fact, the excess in DBS mutation burden observed in eight cancer types (breast, oesophageal, stomach, cervical, ovarian serous and lung squamous carcinomas, cholangiocarcinoma and lung adenocarcinoma) was uniquely linked to platinum treatment mutations (Extended Data Fig. 2b, top bars). Likewise, median mutation contribution from the radiotherapy ID signature[24] (ID8) was enriched in six cancer types commonly exposed to radiation-based treatment (Extended Data Fig. 2c), whereas the 5-fluorouracil[25] (SBS17a/b) and polycyclic aromatic hydrocarbon metabolites from chemotreatments[26] (DBS2) also displayed greater metastatic mutation burden contribution in a tumour-type-specific manner (Extended Data Fig. 2b,c).

The broad enrichment of SBS2/SBS13 mutations in metastatic cancers suggests enhanced activity of APOBEC mutagenesis during the progression of advanced tumours. Specifically, our results revealed an increase in APOBEC mutation burden of 325 ± 178 (mean ± s.d.) mutations per sample in six metastatic tumours (breast, colorectal, stomach, kidney, prostate and pancreatic neuroendocrine carcinomas) that reached statistical significance, with breast and stomach cancers the types with the strongest increase (more than 500 APOBEC mutations per sample). Other cancer types, such as cervical and bladder urothelial carcinomas, also showed enhanced APOBEC activity (more than 2,500 median mutations per sample), but they did not

reach significance due to already high intrinsic APOBEC activity in the primary tumours. The metastatic breast cancer samples also had a higher percentage in clustered APOBEC hypermutation variants than primary tumours (15% versus 5%; Extended Data Fig. 2d and Supplementary Table 3).

Six metastatic cancer types also displayed more mutations from the clock-like mutational processes, including five cancer types (diffuse large B cell lymphoma, breast, prostate, pancreatic neuroendocrine and kidney renal clear cell carcinomas) that exhibited an increased SBS1 contribution and three cancer types (hepatocellular, prostate and thyroid carcinomas) that had an increased SBS5/SBS40 mutation burden. The increase in clock-like mutations in thyroid and prostate cancers, as well as diffuse B cell lymphomas to a lesser extent, may be explained by a larger proportion of older patients with metastatic disease. However, SBS1 metastatic enrichment was also present in cancer types with highly similar age population distributions (Fig. 1a).

Additional focused analyses will be needed to obtain a better understanding of the mutational signature differences that we observed in smaller subsets of cancer types (and subtypes) and metastatic locations (full data in Supplementary Table 3 and Supplementary Data 1).

Differential SBS1 mutation burden

To investigate the SBS1 mutation burden differences in more detail, we evaluated their SBS1 mutation burden by the age of biopsy separately for both cohorts. As expected, the SBS1 mutation burden per year was highly tissue specific[27,28] and displayed an increase with age in the majority of cancer types in both primary and metastatic cohorts (Pearson's $R > 0.1$, 15 of 23 tumour types; Extended Data Fig. 4a and Supplementary Table 4). However, four cancer types (that is, breast, prostate, kidney renal clear cell and thyroid carcinomas) showed an age-independent and significant enrichment of SBS1 mutations in metastatic lesions (Extended Data Fig. 4a). For instance, metastatic breast cancer had a nearly uniform fold-change increase of 1.46 over primary tumours (188 ± 16 SBS1 mutations, mean ± s.d.) across the ages of biopsies, and which was generally consistent across

breast cancer subtypes (Extended Data Fig. 4b and Supplementary Note 3). This pattern was highly cancer-type specific and was not observed for most cancer types, including those with similar intratumour heterogeneity in the primary cohort (for example, colorectal, ovarian serous and pancreatic carcinomas) (Extended Data Fig. 4a). Moreover, this pattern was not explained by differences in tumour genome ploidy (Extended Data Fig. 4c) or by metastatic biopsy location (Supplementary Note 3), was observed in paired primary–metastatic biopsies from individual patients with breast and kidney renal clear cell carcinomas and rendered consistent patterns when relying on independent unpaired cohorts (Supplementary Note 3). Finally, other mutational processes that operate over the evolution of the somatic tissues (for example, clock-like mutations attributed to SBS5/SBS40 that accumulate with age in a cell-cycle-independent manner[28,29]) did not show such enrichment (Extended Data Fig. 4d).

SBS1 mutation burden has been extensively correlated with estimated stem cell division rates[30]. Therefore, an increase in age and tumour-type-specific SBS1 mutation burden in treated metastatic tumours may indicate that these tumours have undergone a higher number of cell divisions. However, the estimated number of years to explain the SBS1 mutation burden shift (23 and 71 years for breast and prostate cancers, respectively; see Supplementary Table 4) shows that this is unlikely to be the main cause. Hence, a more plausible explanation, which also supports previous observations[31,32], is that these metastatic tumours display accelerated cell division rates compared with their primary tumour counterparts (Extended Data Fig. 4e). Supporting this hypothesis, metastatic tumours also had a lower normalized fraction of clonal SBS1 mutations (Extended Data Fig. 4f). Of note, this pattern was not observed in cancer types with consistently high SBS1 mutagenic dynamics (Extended Data Fig. 4f) and was indistinguishable for SBS5/SBS40 mutations (Supplementary Note 3).

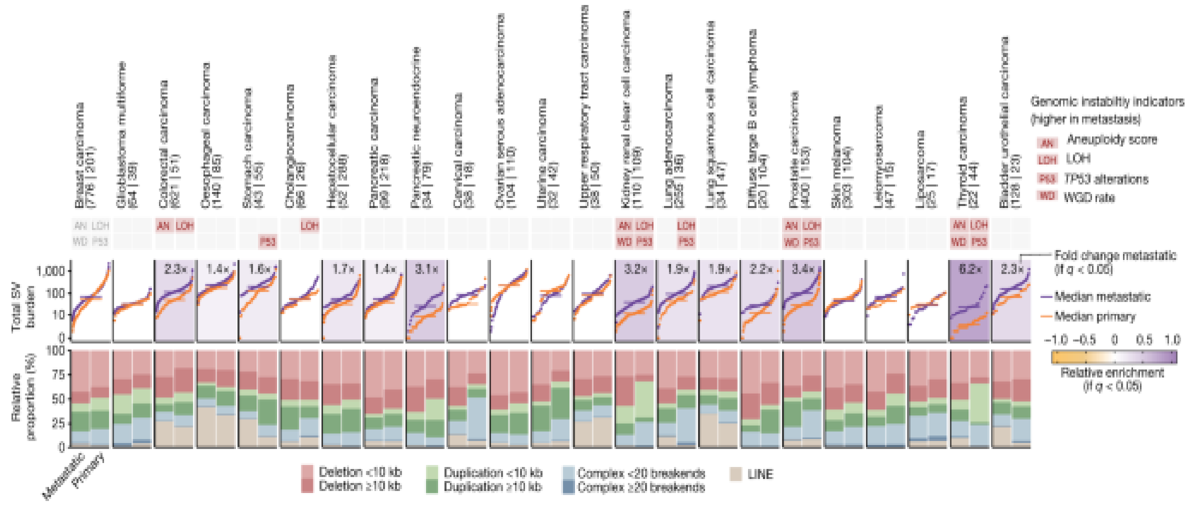
Finally, we observed a negative association between the yearly rate of SBS1 mutation accumulation in primary tumours (a proxy of stem cell division rates[30]) and the estimated fold change of the SBS1 mutation rate in the metastatic cohort (Extended Data Fig. 4g,h). This suggests that tumours with an intrinsically active turnover rate (for example, colorectal carcinomas) preserve their high proliferation rates, whereas others with lower cell

division rates (for example, breast, prostate, kidney and thyroid carcinomas) may acquire higher proliferation rates during the course of cancer progression. Nevertheless, we cannot rule out the contribution of other tissue-type-specific or tumour-type-specific mechanisms, such as higher rates of 5-methylcytosine deamination, decreased fidelity to repair these mismatches or higher contribution from other metastatic-specific mutational processes with overlapping mutational contexts.

SV burden

Comparison of the total number of SVs per tumour revealed an extensive increase in the metastatic tumours (fold change of 2.5 ± 1.3, mean ± s.d.). This increase was observed in 13 of 23 (56%) cancer types (Fig. 3a, Extended Data Fig. 5a and Supplementary Table 5), and was not generally explained by differences in sequencing coverage, tumour clonality (Supplementary Note 1) or cancer-subtype composition (Supplementary Note 2). Moreover, the increased SV burden was also observed for cancer types lacking substantial changes in genomic instability indicators, such as oesophageal and lung squamous cell carcinomas (Fig. 3a). Finally, we observed an increased SV burden in prostate and pancreatic neuroendocrine primary tumours that eventually progressed compared with those with relatively better prognosis, which in both cases were in turn lower than the median values in metastatic tumours (Supplementary Note 2). Overall, compared with TMB, the SV analyses revealed a much more widespread pan-cancer effect, with larger increases per metastatic cancer type that affected almost every cancer type studied.

Fig. 3: SV burden.



The top rectangles represent the four genomic instability features defined in Fig. 1e. The red background represents significant enrichment in the metastatic cohort (two-sided Mann–Whitney adjusted $P < 0.01$). S-plots and cumulative distribution function plots (samples ranked independently for each SV type) of the aggregated SV burden for each cancer type. The horizontal lines represent median values. Backgrounds are coloured according to the relative enrichment, defined as: $\log_{10}$(median SV-type burden in metastatic tumours + 1) − $\log_{10}$(median SV-type burden in primary tumours + 1). Fold-change labels and coloured backgrounds are displayed when Mann–Whitney comparison renders a significant $q < 0.05$. Fold-change labels are displayed with '>' when the SV burden for primary tumours is 0 (see Methods for more details). For each cancer type, the bottom bar plots represent the relative fraction of each SV type in the metastatic (left) and primary (right) datasets. LINE, long interspersed nuclear element.

Small (less than 10 kb) deletions were the most enriched SV types in metastatic tumours (2.7 ± 1.2 fold change in 15 of 23 cancer types with significant enrichment, mean ± s.d.; Extended Data Fig. 5a,b). Larger (10 kb or larger) deletions and duplications had a similar pan-cancer enrichment, although generally with slightly lower fold changes. Complex SVs with 20 or more breakpoints, encompassing events such as chromothripsis and chromoplexy, were enriched in metastatic prostate carcinomas (more than threefold enrichment). Finally, a strong cancer-type-specific metastatic enrichment was also noted for long interspersed nuclear element insertions

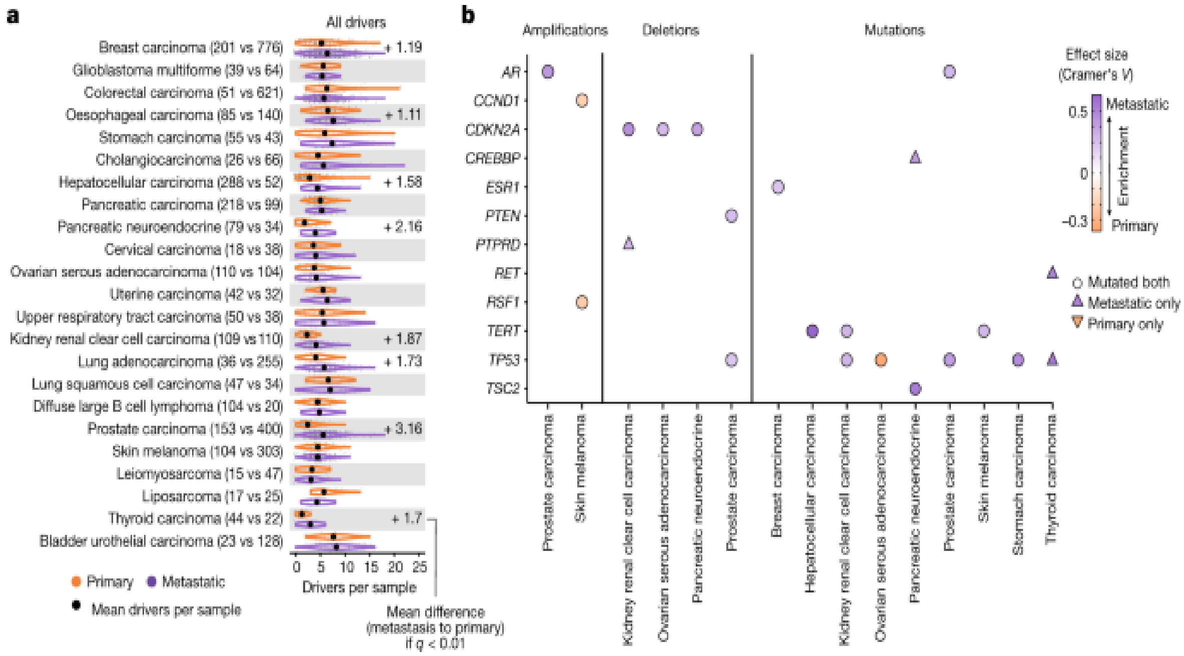
(LINE), with an increased fold change of 12.2 and 12.5 in stomach and bladder urothelial carcinomas, respectively.

We next used linear regression models to unravel the underlying features associated with the observed increase in SV burden in metastatic tumours (Extended Data Fig. 6a–k and Supplementary Table 5). Our approach confirmed the role of previously described cancer-type-specific driver-induced SV phenotypes, including homologous recombination deficiency[33] in metastatic breast carcinoma tumours, *CDK12* (ref. [34]) alterations in prostate carcinoma and *MDM2* (ref. [35]) amplifications in breast $ER^+/HER2^-$ carcinomas, among others. Genomic instability features, such as genome ploidy and *TP53* alterations, showed a strong pan-cancer association with deletions and duplications (Extended Data Fig. 6a,d), and thus very likely contributed to the observed SV increase in metastatic tumours[20,22,23]. Finally, previous exposure to radiotherapy treatment was strongly associated with small deletions in breast $ER^+/HER2^-$ and prostate carcinomas[36].

Cancer driver gene landscape

Metastatic tumours showed a moderate increase in the total number of driver gene alterations per patient (a mean of 4.5 and 5.3 driver alterations per sample in primary and metastatic tumours, respectively), including 8 (34%) tumour types with a significant increase (Fig. 4a). Prostate carcinoma (average increase of 3.16 driver alterations per sample), pancreatic neuroendocrine tumour (2.16), thyroid carcinoma (1.7) and kidney renal clear cell carcinoma (1.87) showed the strongest increases (more than 1.5 driver alterations per patient), whereas the majority of cancer types showed a mean increase below 1.5 driver alterations per sample. All mutation types (amplifications, deletions and mutations) contributed to the increased driver alterations in metastatic tumours (Extended Data Fig. 7a).

Fig. 4: Driver alterations in primary and metastatic tumours.



a, Cancer-type-specific distribution of the number of driver alterations per patient in primary (top) and metastatic (bottom) tumours. The black dots represent the mean values. Labels display mean differences (metastatic to primary) in cancer types with a significant difference (two-sided Mann–Whitney adjusted $P < 0.01$). **b**, Heatmap representing the cancer genes displaying significant mutational frequency differences between primary and metastatic tumours (two-sided Fisher's exact test adjusted $P < 0.01$). Circles denote mutation frequency enrichment in both cohorts, whereas triangles facing upwards and downwards represent drivers that are exclusively enriched in metastatic and primary cohorts, respectively. Colours represent the direction of the enrichment.

Comparison of gene and cancer-type frequencies revealed that only 12 genes had a significant frequency bias in at least one cancer type (22 gene and cancer-type pairs in total; Fig. 4b, Extended Data Fig. 7b and Supplementary Table 6). The majority (19 out of 22, 86%) of the significant pairs had enrichment towards higher metastatic frequency, including four driver genes that were exclusively mutated in metastatic tumours and were not found in the primary tumour equivalents (*PTPRD* in kidney renal clear cell carcinoma, *CREBBP* in pancreatic neuroendocrine tumour, and *RET* and *TP53* alterations in thyroid carcinoma). Most metastatic-enriched cancer drivers had a cancer-type-specific enrichment, including well-established resistance gene drivers associated with anticancer therapies, such as *AR* and

ESR1 alterations in patients with prostate and ER^+ breast carcinomas treated with hormone deprivation therapies[37,38] (Fig. 4b and Supplementary Note 2). Nevertheless, three driver genes (that is, *TP53*, *CDKN2A* and *TERT*) showed a metastatic enrichment across multiple cancer types (Fig. 4b), indicating that alterations of these genes may enhance aggressiveness by disturbing pan-cancer hallmarks of tumorigenesis.

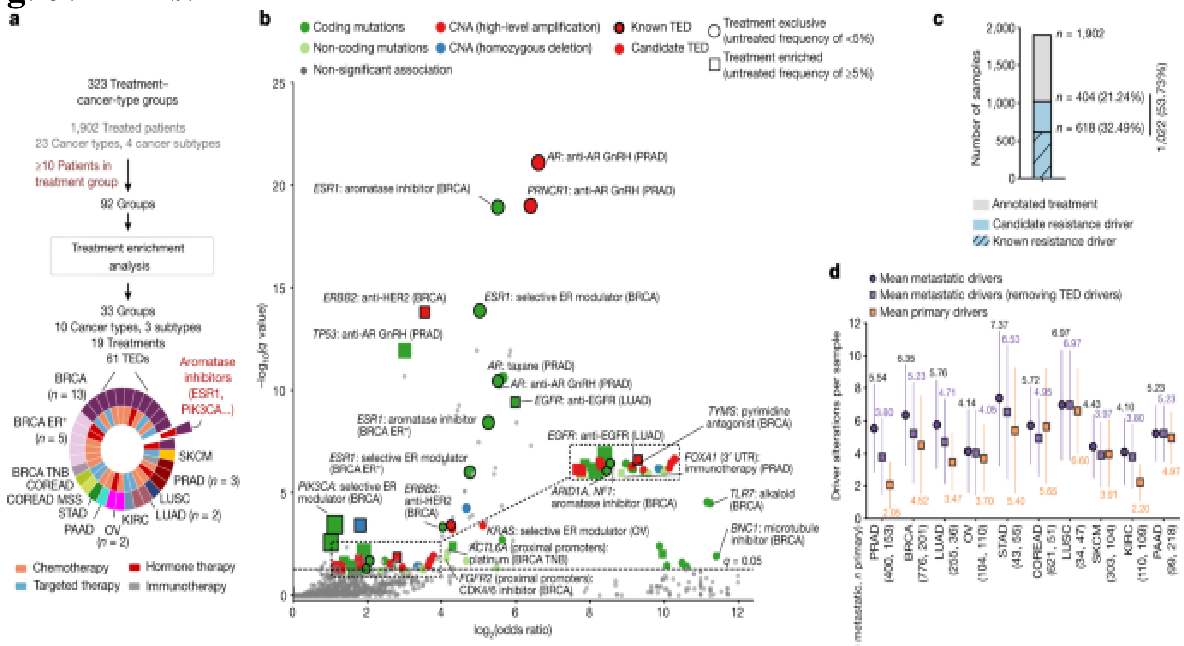
We next investigated whether the reported driver differences may have an effect on potential clinical actionability. Cancer-type-specific comparison of therapeutically actionable variants revealed an overall larger fraction of patients with therapeutically actionable variants in the metastatic cohort, with high variability across cancer types (Extended Data Fig. 8a and Supplementary Table 7). Subsetting by A-on label variants (that is, approved biomarkers in the specific cancer type) revealed a consistent pattern in which only cholangiocarcinoma (*FGFR2* fusions and *IDH1* mutations) and lung adenocarcinoma (*EGFR* alterations) showed a substantial proportional increase in the metastatic cohort (Extended Data Fig. 8a,b). Non-A-on label biomarkers (A-off label, B-on label and B-off label) showed a modest and tumour-type-dependent metastatic increase, which was mainly linked to the increased alteration frequency of *KRAS* exon 2 mutations and *CDKN2A* loss in advanced tumour stages (Extended Data Fig. 8b).

Treatment-associated drivers

The presence of treatment resistance driver genes in late-stage tumours prompted us to devise a test that aimed to identify treatment-enriched drivers (TEDs) that were either significantly enriched (that is, treatment enriched) or exclusively found (that is, treatment exclusive) in a cancer-type-specific and treatment-specific manner (Extended Data Fig. 9a and Methods). Our analytical framework provided 61 TEDs associated with 33 treatment groups from 8 cancer types and 4 cancer subtypes (Fig. 5a,b, Supplementary Table 8 and Supplementary Note 4). Of the identified TEDs, 33 of 61 (54%) were coding mutation drivers, 16 (26%) were copy number amplifications, 9 (14%) were non-coding drivers and 3 (6%) were recurrent homozygous deletions (Fig. 5b and Extended Data Fig. 9b,c). Reassuringly, the majority of the top hits were known treatment resistance drivers, including *AR*-activating mutations and gene amplifications in patients with prostate cancer

treated with androgen-deprivation therapy[38] (Extended Data Fig. 9d,e), *ESR1*$^{536–538}$ mutations in patients with breast cancer treated with aromatase inhibitors[37] (Extended Data Fig. 9f), and *EGFR*^{T790M} mutations (Extended Data Fig. 9g) and *EGFR* copy number gains in patients with lung adenocarcinoma treated with *EGFR* inhibitors[39,40] (Extended Data Fig. 9h), among others. Moreover, we also found that *TP53* alterations were recurrently associated with resistance to multiple treatments, which may indicate that these alterations are prognostic markers for enhanced tumour aggressiveness and plasticity rather than being a cancer-type-specific mechanism of drug resistance (Supplementary Note 4).

Fig. 5: TEDs.



a, Workflow representing the number of treatment groups in each step of the analysis. For the external layers of the pie chart, the number of treatments with identified TEDs is coloured by cancer type. For the internal layers of the pie chart, the category of the corresponding treatment is shown. *n* Refers to the number of treatment groups with at least one TED. MSS, microsatellite stable; TNB, triple negative breast cancer. **b**, Volcano plots displaying the identified TEDs. Each dot represents one cancer gene alteration type in one treatment group. The *x* axis displays the effect size (as $\log_2$(odds ratio)) and the *y* axis shows the significance ($-\log_{10}$(*q* value)). The circle markers denote TEDs exclusively mutated in the treatment group

(squared makers are used otherwise). Markers are coloured according to the type of alteration. The thicker edge lines indicate known resistance drivers. CNA, copy number alteration; UTR, untranslated region. **c**, Global proportion of patients with TEDs treated for metastasis. **d**, Mean number of driver alterations per patient with a metastatic tumour before (purple circle) and after (purple square) excluding TEDs compared with patients with primary tumours (orange square). The vertical lines indicate s.d. The mean number of driver alterations are labelled. *n* Metastatic and *n* primary denote the number of metastatic and primary samples, respectively.

Our results also provided a long tail of candidate drivers of resistance, some of them with orthogonal evidence by independent reports (Fig. 5b and Extended Data Fig. 9b,c). Examples of the latter group include *TYMS* amplification in patients with breast cancer treated with pyrimidine antagonists[41] (Extended Data Fig. 9i), *PRNC1* and *MYC* co-amplifications in patients with prostate cancer treated with androgen deprivation[42] (Extended Data Fig. 9j), *ACTL6A* promoter mutations in patients with triple-negative breast cancer treated with platinum-based therapies[43], and *FGFR2* promoter mutations in patients with breast cancer treated with CDK4/CDK6 inhibitors[44]. The full TEDs catalogue is provided in Supplementary Table 8 and constitutes a valuable resource for investigating resistance mechanisms to common cancer therapies.

Overall, 53% of patients with metastatic disease with annotated treatment information had TEDs, including 32% with annotations of known resistance drivers and an additional 21% of patients with candidate resistance drivers derived from our analysis (Fig. 5c). We identified 0.70 ± 0.53 (mean ± s.d.) TEDs per metastatic sample across the 8 cancer types that had reported TEDs (Fig. 5d), with prostate and breast carcinomas displaying the greatest prevalence of TEDs (that is, 1.74 and 1.12 drivers per patient with prostate and breast cancers, respectively). Therefore, after excluding TEDs, primary and metastatic tumours had a 36% reduction of their original differences in the number of drivers per sample (from 5.3 to 5.0 mean drivers per sample in the metastatic cohort after excluding TEDs, compared with 4.5 mean drivers per sample in the primary cohort) (Fig. 5d and Supplementary Table 8), indicating that an important proportion of the metastatic-enriched drivers are probably associated with resistance to anticancer therapies.

Discussion

In this study, we describe a cohort of more than 7,000 uniformly reprocessed WGS samples from patients with primary untreated and metastatic treated tumours. We compared genomic features across 23 cancer types and confirmed previous cancer-type-specific observations while also providing novel biological findings, such as the clock-like molecular features, the prevalence of SV burden across different tumorigenic stages and the incidence of TEDs in treated patients.

Specifically, metastatic tumours displayed high genomic instability, low intratumour heterogeneity and strong accumulation of SVs. However, the magnitude of genomic differences between primary and metastatic tumours was highly cancer-type specific and was influenced by the exposure to cancer treatments. Overall, five cancer types (prostate, thyroid, kidney renal clear cell, breast and pancreatic neuroendocrine carcinomas) showed an intense transformation of the genomic landscape in advanced tumorigenic stages (Extended Data Fig. 10, labelled as strong). The other cancer types displayed variable genomic differences, although the chromosomal genomic portrait tended to be conserved.

Cancer types with the strongest genomic differences between primary and metastatic settings in our analyses typically have a good prognosis in the primary setting. But then, whether the metastatic tumours representing a unique set of primary patients that eventually progressed (that is primaries from the metastatic cohort were ‘born to be bad’) or whether there are stochastic triggers of metastatic disease in relatively indolent primaries are still to be determined. To fully address these, larger pan-cancer sets of matched biopsies from the same patient, as already implemented in various cancer-type-specific studies[45,46], would be needed.

This study faced various limitations, such as the use of different laboratory workups and sequencing parameters used for primary and metastatic tumour samples, although we demonstrated that this does not severely have an effect on the overall detectability of clonal somatic variants (Supplementary Note 1). However, we cannot exclude the possibility of missing highly subclonal driver mutations. Furthermore, our observations are unlikely to be

exhaustive, especially for lower frequency events, because of limited cohort (or subcohort) sizes. Expanding cancer cohorts in research or clinical settings will be essential to advance our understanding of tumour progression. Finally, genomic changes alone cannot entirely explain how tumour cells are able to colonize other organs. Therefore, additional information from complementary tumour omics[47] and from the tumour microenvironment[48] will be needed to further dissect and better understand metastasis and resistance to cancer therapies.

To conclude, our dataset constitutes a valuable resource that can be leveraged to further study other aspects of tumour evolution, such as genomic differences across metastatic biopsy locations (Supplementary Note 2), dedicated analysis for cancer subtypes (Supplementary Note 2), genetic immune escape alterations in primary and metastatic tumours[49] as well as for the development of machine learning tools to foster cancer diagnostics[50].

Methods

Cohort gathering and processing

We have matched tumour–normal WGS data from patients with cancer from two independent cohorts: Hartwig and PCAWG. A detailed description of the Hartwig and PCAWG cohort gathering and processing as well as comprehensive documentation of the PCAWG sample reanalysis with the Hartwig somatic pipeline is described in the Supplementary Note 1.

Tumour clonality analysis

Each mutation in the .vcf files was given a subclonal likelihood by PURPLE. Following PURPLE guidelines, we considered mutations with subclonal scores equal or higher than 0.8 to be subclonal and mutations below the 0.8 threshold to be clonal. For each sample, we then computed the average proportion of clonal mutations by dividing the number of clonal mutations by the total mutation burden (including SBS, multinucleotide variants and IDs). Finally, for each cancer type, we used Mann–Whitney test

to assess the significance of the clonality difference between the primary and metastatic tumours. P values were adjusted for false discovery rate (FDR) using the Benjamini–Hochberg procedure. An adjusted $P < 0.05$ was deemed as significant.

In addition, we leveraged biopsy site data in patient reports to further investigate differences in metastatic tumour clonality according to the metastatic biopsy site (see also Supplementary Note 2). If the metastatic biopsy site was in the same organ or tissue as the primary tumour, we considered them as 'local', whereas if the metastatic biopsy site was reported in the lymphoid system or other organs or tissues, they were dubbed as 'lymph' and 'distant', respectively. Cancer types for which there was a minimum of five samples available for each of the biopsy groups were selected and Mann–Whitney test was used to compare the clonality between the biopsy groups.

Karyotype

Chromosome arm level and genome ploidy was estimated as previously described[20].

First, for each chromosome arm, tumour purity and ploidy-adjusted copy number (CN) segments (as determined by PURPLE) were rounded to the nearest integer. Second, arm coverage of each integer CN was calculated as the fraction of chromosome arm bases with the specific CN divided by the chromosome arm length (for example, 60% of all chromosome 5p segments have a CN of 2, 30% have a CN of 1 and 10% have a CN of 3). We defined the arm-level ploidy level as the CN with the highest coverage across the whole arm (in the example above it would be 2). Third, we computed the most recurrent chromosome arm ploidy levels across all chromosome arms per sample (that is, observed genome ploidy).

Next, we estimated the true genome ploidy by taking WGD status (given by PURPLE) into account. If a sample did not undergo WGD, its total expected genome ploidy was deemed to be $2n$. If a sample did undergo WGD and its observed genome ploidy was less than six, the estimated genome ploidy was deemed to be $4n$, and $8n$ if the observed genome ploidy was six or more. An observed genome CN of more than eight was not found in our dataset.

Then, for each chromosome arm in each sample, we defined the normalized arm ploidy as the difference between the arm-level ploidy level and the expected genome ploidy. The resulting value was classified as 1 for differences higher than or equal to 1 (representing arm gains), as −1 for differences lower than or equal to −1 (representing arm losses) or as 0 (no difference). Normalized arm ploidy values were averaged across all samples from a cancer type in a cohort-specific manner (that is, separating primary and metastatic samples). A Mann–Whitney test was performed per cancer type and chromosome arm to assess the mean difference in arm gains or losses at the cancer-type level. The resulting *P* value was FDR adjusted across all arms per cancer type. Finally, $q < 0.01$ and a normalized arm ploidy difference higher than 0.25 was deemed to be significant.

Genomic instability indicators

To compare the differences in aneuploidy scores and the LOH proportions in each group, a Mann–Whitney test was performed per cancer type. The aneuploidy score represents the number of arms per tumour sample that deviate from the estimated genome ploidy as previously described[20]. The LOH score of a given sample represents the sum of all LOH regions divided by the GRCh37 total genome length. A genomic region is defined as LOH when the minor allele CN < 0.25 and major allele CN ≥ 0.8.

To compare the fraction of samples with a driver mutation in *TP53* as well as the fraction of WGD samples per cohort, a Fisher's exact test was performed per cancer type. Any *TP53* driver alteration (non-synonymous mutation, biallelic deletion and homozygous disruption) was considered in the analysis. Multiple driver mutations per sample in a single gene were considered as one driver event. WGD was defined as present if the sample had more than 10 autosomes with an estimated chromosome CN of more than 1.5. *P* values were FDR corrected across all cancer types. A $q < 0.01$ was deemed to be significant for all statistical tests.

Mutational signature analysis

Signature extraction

The number of somatic mutations falling into the 96 SBS, 78 DBS and 83 ID contexts (as described in the COSMIC catalogue[51]; https://cancer.sanger.ac.uk/signatures/) was determined using the R package mutSigExtractor (https://github.com/UMCUGenetics/mutSigExtractor, v1.23).

SigProfilerExtractor (v1.1.1) was then used (with default settings) to extract a maximum of 21 SBS, 8 DBS and 10 ID de novo mutational signatures. This was performed separately for each of the 20 tissue types that had at least 30 patients in the entire dataset (aggregating primary and metastatic samples; see Supplementary Table 3). Tissue types with less than 30 patients as well as patients with metastatic tumours with unknown primary location type were combined into an additional ‘Other’ group, resulting in a total of 21 tissue-type groups for signature extraction. To select the optimum rank (that is, the eventual number of signatures) for each tissue type and mutation type, we manually inspected the average stability and mean sample cosine similarity plot output by SigProfilerExtractor. This resulted in 440 de novo signature profiles extracted across the 21 tissue-type groups (Supplementary Table 3). Least squares fitting was then performed (using the fitToSignatures() function from mutSigExtractor) to determine the per-sample contributions to each tissue-type-specific de novo signature.

Aetiology assignment

The extracted de novo mutational signatures with high cosine similarity (≥0.85) to any reference COSMIC mutational signatures with known cancer-type associations[51] were labelled accordingly (288 de novo signatures matched to 57 COSMIC reference signatures).

For the remaining 152 unlabelled de novo signatures, we reasoned that there could be one or more signatures from one cancer type that is highly similar to those found in other tissue types, and that these probably represent the same underlying mutational process. We therefore performed clustering to group likely equivalent signatures. Specifically, the following steps were performed:

1. (1)

We calculated the pairwise cosine distance between each of the de novo signature profiles.

2. (2)

   We performed hierarchical clustering and used the base R function cutree() to group signature profiles over the range of all possible cluster sizes (minimum number of clusters = 2; maximum number of clusters = number of signature profiles for the respective mutation type).

3. (3)

   We calculated the silhouette score at each cluster size to determine the optimum number of clusters.

4. (4)

   We grouped the signature profiles according to the optimum number of clusters. This yielded 27 SBS, 7 DBS and 8 ID de novo signature clusters (see Supplementary Table 3).

For certain de novo signature clusters, we could manually assign the potential aetiology based on their resemblance to signatures with known aetiology described in COSMIC[51], Kucab et al.[26] and Signal (access date 1 February 2023)[52]. Some clusters were an aggregate of two known signatures, such as SBS_denovo_clust_2, which was a combination of SBS2 and SBS13, both linked to APOBEC mutagenesis. Other clusters had characteristic peaks of known signatures, such as DBS_denovo_clust_4, which resembled DBS5 based on having distinctive CT>AA and CT>AC peaks. Finally, DBS_denovo_clust_1 was annotated as a suspected *POLE* mutation and MMRd, as samples with high contribution (more than 150 mutations) of this cluster are frequently microsatellite instable (MSI) or have *POLE* mutations. Likewise, DBS_denovo_clust_2 was annotated as a suspected MMRd as the aetiology, as samples with high contribution (more than 250 mutations) of this cluster were all MSI. See Supplementary Table 3 for a list of all the manually assigned aetiologies.

After applying the aetiology assignment, the de novo extraction resulted in 69 SBS, 13 DBS and 18 ID representative mutational signatures (Supplementary Table 3). Most of these (42 of 69 SBSs, 7 of 13 DBSs and 8 of 18 IDs) mapped onto the well-described mutational signatures in human cancer[35,53].

Comparing the prevalence of mutational processes between primary and metastatic cancer

We then compared the activity (that is, the number of mutations contributing to) of each mutational process between primary and metastatic tumours. For each sample, we first summed the contributions of signatures of the same mutation type (that is, SBS, DBS or ID) with the same aetiology, hereafter referred to as 'aetiology contribution'. Per cancer type and per aetiology, we performed two-sided Mann–Whitney tests to determine whether there was a significant difference in aetiology contribution of primary and metastatic tumours. Per cancer type and per mutation type, we used the p.adjust() base R function to perform multiple testing correction using Holm's method. Next, we added a pseudocount of 1 to the contributions (to avoid dividing by 0) and calculated the median contribution $\log_2$ fold change, that is, $\log_2$((median contribution in metastatic tumours + 1)/(median contribution in primary tumours + 1)). We considered the aetiology contribution between primary and metastatic tumours to be significantly different when $q < 0.05$, and $\log_2$ fold change ≥ 0.4 or $\log_2$ fold change ≤ -0.4 ($= \pm \times 1.4$).

Relative contribution

Relative aetiology contribution was calculated by dividing aetiology contribution by the total contribution of the respective mutation type (that is, SBS, DBS or ID). To determine the significant difference in relative aetiology contribution, we performed two-sided Mann–Whitney tests as described above. We also calculated the median difference in contribution (that is, median relative contribution in metastatic tumours − median relative contribution in primary tumours). We considered the relative aetiology contribution between primary and metastatic tumours to be significantly different when $q < 0.05$ and median difference was 0.01 or more.

We also determined whether there was an increase in the number of samples with high aetiology contribution (that is, hypermutators) in metastatic versus primary cohorts. For each signature, a sample was considered a hypermutator if the aetiology contribution was 10,000 or more for SBS signatures, 500 or more for DBS signatures or 1,000 or more for ID signatures. For each cancer type, for each aetiology, we performed pairwise testing only for cases in which there were five or more hypermutator samples for either metastatic or primary tumours. Each pairwise test involved calculating *P* values using two-sided Fisher's exact tests, and effect sizes by multiplying Cramer's *V* by the sign of the $\log_2$(odds ratio) to calculate a signed Cramer's *V* value that ranges from −1 to +1 (indicating enrichment in primary or metastatic, respectively). We then used the p.adjust() base R function to perform multiple testing correction using Bonferroni's method.

SBS1–age correlations in primary and metastatic tumours

To count the SBS1 mutations, we relied on the definition from ref. [54] that is based on the characteristic peaks of the COSMIC SBS1 signature profile: single-base CpG > TpG mutations in NpCpG context. To ensure that these counts and the downstream analyses are not affected by differential APOBEC exposure in primary and metastatic cohorts, we excluded CpG > TpG in TpCpG, which is also a characteristic peak in the COSMIC SBS2 signature profile. In addition, for skin melanoma, CpG > TpG in [C/T]pCpG, which overlaps with SBS7a, was excluded. To obtain the SBS5 and SBS40 counts, we relied on their exposures derived from the mutational signature analyses performed in this study (described above).

To assess the correlation between SBS1 burden and the age of the patient, at biopsy we performed a cancer-type and cohort-specific linear regression (that is, separate regression for primary and metastatic tumour samples). To avoid spurious effects caused by hypermutated tumours, samples with a TMB greater than 30,000 as well as those with SBS1 burden greater than 5,000 were excluded.

For each cancer type and cohort, we then computed 100 independent linear regressions by randomly selecting 75% of the available samples. We

selected the median linear regression (based on the regression slope) as representative regression for further analyses. Similarly, confidence intervals were derived from the 1st and 99th percentile of the computed regressions.

To evaluate the significance of the differences between primary and metastatic representative linear regressions (hereafter referred to as linear regression for simplicity), we first filtered out cancer types that failed to show a positive correlation trend between SBS1 burden and age at biopsy in both primary and metastatic tumours (that is, Pearson's correlation coefficient of primary and metastatic regression greater than 0.1). Next, for each selected cancer type, we computed the regression residuals of primary and metastatic SBS1 mutation counts using, in both cases, the primary linear regression as baseline. The primary and metastatic residual distributions were then compared using a Mann–Whitney test to evaluate significance. Cancer types with a Mann–Whitney $P < 0.01$ were deemed as significant. Finally, to ensure that the differences were uniform across different age ranges (that is, not driven by a small subset of patients), we only considered significant cancer types in which the metastatic linear regression intercept was higher than the primary intercept.

SBS5/SBS40 correlations were computed following the same procedure and using the sum of SBS5 and SBS40 exposures for each tumour sample. If none of the mutations were attributed to SBS5/SBS40 mutational signatures, the aggregated value was set to zero. In the ploidy-corrected analyses, we divided the SBS1 mutation counts (and SBS5/SBS40 mutation counts for the SBS5/SBS40 ploidy-corrected regression, respectively) by the PURPLE-estimated tumour genome ploidy.

For each cancer type, the mean fold change (fc) was defined as $\underline{{\rm{fc}}}=\frac{1}{40}{\sum }_{i=40}^{80}\frac{{{\rm{M}}{\rm{Pred}}}_{i}}{{{\rm{P}}{\rm{Pred}}}_{i}}$ where $MPred_i$ and $PPred_i$ are the estimated number of SBS1 mutations for a given age ith according to the metastatic and primary linear regressions, respectively. Similarly, the mean estimated SBS1 burden difference ($SBS1_{diff}$) was defined as: $\underline{{\rm{SBS1diff}}}=\frac{1}{40}{\sum }_{i=40}^{80}{{\rm{MPred}}}_{i}-{{\rm{PPred}}}_{i}$.

Clonality of clock-like mutations

SBS1 individual mutations were identified as described in the previous section. For SBS5 and SBS40 mutations, we used a maximum likelihood approach to assign individual mutations to the SBS5 and SBS40 mutational signatures in a cancer-type-specific manner.

For every SBS1 (and SBS5/SBS40 mutation), we then assign the clonality according to the PURPLE subclonal likelihood estimation, in which only mutations with subclonal (SUBCL) likelihood ≥ 0.8 were considered as such (see above).

For each tumour sample, the SBS1 clonality ratio (or respectively SBS5/SBS40 clonality ratio) was defined as the ratio between the proportion of clonal SBS1 mutations ($\frac{{\rm{SBS\; 1\; clonal\; mutations}}}{{\rm{SBS\; 1\; mutations}}}$) divided by the total proportion of clonal alterations in the sample ($\frac{{\rm{Total\; clonal\; mutations}}}{{\rm{Total\; mutations}}}$).

Primary SBS1 mutation rate and metastatic SBS1 age-corrected enrichment

We computed for each primary cancer type the average number of SBS1 per year as the number of SBS1 mutations divided by the age of the patient at biopsy (only considering primary samples and excluding hypermutated samples as described above). We then used a Spearman's correlation to assess its association with the estimated mean SBS1 mutation rate fold change in metastatic tumours (see above). In addition, to exclude potential biases in our primary cohort, we repeated the same analysis relying on an independent measurement of primary cancer SBS1 yearly accumulation. Specifically, we used the best-estimated accumulation of SBS1 per year from ref. [30] (Supplementary Table 6) and regressed it to the fold-change estimates for the matching cancer types present in both datasets.

SV analysis

Definitions of SV type

LINX[55] chains one or more SVs and classifies these SV clusters into various event types ('ResolvedType'). We defined deletions and duplications as clusters with a ResolvedType of 'DEL' or 'DUP' whose start and end breakpoints are on the same chromosome (that is, intrachromosomal). Deletions and duplications were split into those less than 10 kb and 10 kb or more in length (small and large, respectively), based on observing bimodal distributions in these lengths across cancer types (Extended Data Fig. 5b). We defined complex SVs as clusters with a 'COMPLEX' ResolvedType, an inversion ResolvedType (including: INV, FB_INV_PAIR, RECIP_INV, RECIP_INV_DEL_DUP and RECIP_INV_DUPS) or a translocation ResolvedType (including: RECIP_TRANS, RECIP_TRANS_DEL_DUP, RECIP_TRANS_DUPS, UNBAL_TRANS and UNBAL_TRANS_TI). Complex SVs were split into those with less than 20 and 20 or more SVs (small and large, respectively), based on observing similar unimodal distributions in the number SVs across cancer types whose tail begins at approximately 20 breakpoints (Extended Data Fig. 5b). Finally, we defined long interspersed nuclear element insertions (LINEs) as clusters with a ResolvedType of 'LINE'. For each sample, we counted the occurrence (that is, SV burden) of each of the seven SV types described above. In addition, we determined the total SV burden by summing counts of the SV types.

Comparing SV burden between primary versus metastatic cancer

We then compared the SV-type burden between primary versus metastatic tumours as shown in Fig. 3a. First, we performed two-sided Mann–Whitney tests per SV type and per cancer type to determine whether there was a statistically significant difference in SV-type burden between primary versus metastatic. The Bonferroni method was used for multiple testing correction on the *P* values from the Mann–Whitney tests (to obtain *q* values). Next, we calculated relative enrichment as follows: $\log_{10}$(median SV-type burden in metastatic tumours + 1) − $\log_{10}$(median SV-type burden in primary tumours + 1); and calculated fold change as follows: (median SV-type burden in metastatic tumours + 1) / (median SV-type burden in primary tumours + 1). When calculating relative enrichment and fold change, the pseudocount of 1 was added to avoid the log(0) and divide by zero errors, respectively. Fold changes are displayed with a '>' in Fig. 3a when the SV burden for primary tumours is 0 (that is, when a divide by 0 would occur without the

pseudocount). We considered the SV-type burden between primary versus metastatic to be significant when: $q < 0.05$, and fold change ≥ 1.2 or fold change ≤ 0.8

Identifying features associated with increased SV burden in metastatic cancer

To identify the features that could explain increased SV burden, we correlated SV burden with various tumour genomic features. This included: (1) genome ploidy (determined by PURPLE); (2) homologous recombination deficiency (determined by CHORD[33]) and MSI (determined by PURPLE) status; (3) the presence of mutations in 345 cancer-associated genes (excluding fragile site genes that are often affected by CN alterations[5]), hereafter referred to as ‘gene status’; and (4) treatment history, including the presence of radiotherapy, the presence of one of the 79 different cancer therapies as well as the total number of treatments received. All primary samples and all metastatic samples without treatment information were considered to have no treatment. Genome ploidy and total number of treatments received were numeric features, whereas all of the remaining were boolean (that is, true or false) features. In total, there were 429 features.

SV-type burden was transformed to $\log_{10}$(SV-type burden + 1) and was correlated with the 429 features using multivariate linear regression models (LMs). This was performed separately for each of the seven SV types, and for each cancer type (or subtype). In the SV main analysis (Fig. 4b–f), there were 23 cancer types, resulting in a total of 161 (23 cancer types × 7 SV types) LM models.

Each LM model (that is, per SV type and cancer type) involved training of three independent LMs with (1) both metastatic and primary samples (primary + metastatic), (2) only Hartwig samples (metastatic only), and (3) only PCAWG samples (primary only). This was done to filter out correlations between features and increased SV-type burden solely due to differences in feature values between primary and metastatic tumours. We then required features that positively correlated with SV-type burden in the primary + metastatic LM to independently show the same association in the

metastatic-only or primary-only LMs. Only genomic features that independently showed positive correlation with the SV burden were further considered as significant (that is, represented in the lollipop plots).

Each of the three LMs was trained as follows:

1. (1)

   Remove boolean features with too few ‘true’ samples.

1. (i)

   For the primary + metastatic LM, remove gene status features with less than 15 ‘true’ samples.

2. (ii)

   For the metastatic-only and primary-only LMs, remove gene status features with less than 10 ‘true’ samples.

3. (iii)

   For the remaining boolean features, remove features with less than 5% ‘true’ samples.

1. (2)

   Fit a LM using the lm() base R function to correlate $\log_{10}$(SV-type burden + 1) versus all features.

For each LM analysis, we used the following filtering criteria to identify the features that were correlated with increased SV-type burden:

1. (1)

   Only keep LM analyses for which there was significant increase in SV-type burden for the respective cancer type ($P < 0.01$ as described in the previous section ‘Comparing SV burden between primary versus metastatic cancer’).

2. (2)

   For primary + metastatic LM:

1. (i)

   Regression $P < 0.01$

2. (ii)

   Coefficient $P < 0.01$

3. (iii)

   Coefficient more than 0

1. (3)

   For metastatic-only LM or primary-only LM:

1. (i)

   Coefficient $P < 0.01$

2. (ii)

   Coefficient more than 0

Finally, to determine which features (of those correlated with increased SV-type burden) were enriched in metastatic tumours compared with primary tumours (and vice versa), we calculated Cliff's delta for numeric features and Cramer's V for boolean features. Cliff's delta ranges from −1 to +1, with −1 representing complete enrichment in primary tumours, whereas +1 represents complete enrichment in metastatic tumours. Cramer's V only ranges from 0 to 1 (with 1 representing enrichment in either primary or metastatic tumours), the sign of the log(odds ratio) was assigned as the sign of the Cramer's V value so that it ranged from −1 to +1. Features with an effect size of more than 0 were considered as those that could explain the SV burden increase in metastatic cancer when compared with primary cancer.

Driver alterations

We relied on patient-specific cancer driver and fusion catalogues constructed by PURPLE[5] and LINX[55]. Only drivers with a driver likelihood of more than 0.5 were retained. Fusion drivers were filtered for those that were previously reported in the literature. Similarly, we manually curated the list of drivers and removed *SMAD3* hotspot mutations because of the high-burden mutations in low-mappability regions. The final driver catalogue contained a total of 453 driver genes and the final fusion catalogue contained 554 reported fusions.

To compare the number of drivers in primary and metastatic tumours, we then combined fusions with the LINX driver variants to calculate a patient-specific number of driver events. Drivers that concern the same driver gene but a different driver type were deemed to be single drivers (for example, *TP53* mutation and *TP53* deletion in the same sample were considered as one driver event). Cancer-type-specific Mann–Whitney test was performed to assess differences between primary and metastatic tumours. An adjusted $q < 0.01$ was deemed to be significant.

To assess the driver enrichment, a contingency matrix was constructed from the driver catalogue, containing the frequency of driver mutations per driver type (that is, deletion, amplification or mutations) and cancer type in each cohort (metastatic and primary). A second contingency matrix was constructed for the fusions. Partial amplifications were considered as amplifications, whereas homologous disruptions were considered as deletions. These contingency matrices were filtered for genes that show a minimum frequency of five mutated samples in either the primary or the metastatic cohorts. Then, a two-sided Fisher's exact test for each gene, cancer type and mutation type was performed and the P value was adjusted for FDR per cancer type. Cramer's V and the odds ratio were used as effect size measures. An adjusted $P < 0.01$ was deemed to be significant.

Therapeutic actionability of variants

To determine the amount of actionable variants observed in each sample, we compared our variants annotated by SnpEff (v5.1)[56] to those derived from

three different databases (OncoKB[57], CIViC[58] and CGI[59]) that were classified based on a common clinical evidence level (https://civic.readthedocs.io/en/latest/model/evidence/level.html) as previously described[5]. In our study we only considered A and B levels of evidence, which represent variants that have been FDA approved for treatment and are currently being evaluated in a late-stage clinical trial, respectively. A variant was determined to be ‘on-label’ when the cancer type matches the cancer type for which the treatment was approved for or is being investigated for, and ‘off-label’ otherwise. Only actionable variants of the sensitive category were considered (that is, tumours containing the variant are sensitive to a certain treatment). Sample-level actionable variants such as TMB high/low or MSI status were not evaluated, because of their tendency to overshadow the other variants, especially in the off-label category. Furthermore, wild-type actionable variants were not considered in this analysis for the same reason. Variants related to gene expression or methylation were not considered due to lack of available data. In addition, we found actionable variants derived from leukaemias to be very different from the solid tumours in our dataset, which is why we excluded them for this analysis. For the analysis of proportion of samples bearing therapeutically actionable variants, we considered that the highest evidence level was retained for each sample following the order A on/off-label to B on/off-label. To assess enrichment of actionable variants globally and at the A on-label level in metastatic tumours, a Fisher’s exact test was performed pan-cancer-wide and per cancer type. An adjusted $P < 0.05$ was deemed to be significant. Fold changes in frequency are only shown for cancer types with a global significant difference.

To determine which variants contribute the most to the observed significant frequency differences, individual actionable variants were tested for enrichment in metastatic tumours using a Fisher’s exact test per cancer type and tier level. P values were FDR adjusted per cancer type and $q < 0.05$ was deemed to be significant. In Extended Data Fig. 8, only actionable variants from cancer types with a global significant difference (see above) and that were found at a minimum frequency of 5% in either primary or metastatic cohort and a minimum frequency difference of 5% between them were shown. However, the differences across all screened variants are available as part of Supplementary Table 7.

TEDs

We aimed to pinpoint drivers that are potentially responsible for lack of response to certain cancer treatments in the metastatic cohort. Hence, we devised a test that identifies driver alterations that are enriched in groups of patients treated with a particular treatment type compared with the untreated group of patients from the same cancer type (see Extended Data Fig. 9a for illustration of the workflow).

Treatments were grouped according to their mechanism of action so that multiple drugs with a shared mechanism of action were grouped into the mechanistic treatment category (for example, cisplatin, oxaliplatin and carboplatin were grouped as platinum). We created 323 treatment and cancer-type groups by grouping patients with treatment annotation according to their treatment record before the biopsy. One patient might be involved in multiple groups if they have received multiple lines of therapy or a simultaneous combination of multiple drugs. Only 92 treatment and cancer-type groups with at least ten patients were further considered in the analysis.

Hence, for each cancer type (or subtype, in the case of breast and colorectal) and treatment group, we performed the following steps:

1. (1)

   We first performed a driver discovery analysis in treatment and cancer-type (or subtype)-specific manner. We explored three types of somatic alterations: coding mutations, non-coding mutations and CN variants (see below for detailed description of each driver category). Driver elements from each alteration category were selected for further analysis.

2. (2)

   For each driver alteration from (1), we compared the alteration frequency in the treated group to the untreated group of the same cancer type. Each driver category (coding and non-coding mutations and CN variants) were evaluated independently. We performed a Fisher's exact test to assess the significance of the frequency differences. Similarly,

we computed the odds ratio of the mutation frequencies for each driver alteration. The *P* values were adjusted with a multiple-testing correction using the Benjamini–Hochberg procedure ($\alpha = 0.05$). An adjusted *P* value of 0.05 was used for coding mutations and CN variants. An adjusted *P* value of 0.1 was used for non-coding variants due to the overall low mutation frequency of the elements included in this category, which hampered the identification of significant differences.

3. (3)

   We then annotated each driver element with information about the exclusivity in the treatment group. We labelled drivers as treatment exclusive if the mutation frequency in the untreated group was lower than 5% or we annotated as treatment enriched otherwise. In addition, we manually curated the identified drivers with literature references of their association with each treatment category.

4. (4)

   Finally, the overlap of patients in multiple treatment groups (see above) in the same cancer type prompted us to prioritize the most significant treatment association for each driver gene in a particular cancer type. In other words, for each driver gene that was deemed as significantly associated with multiple treatment groups in the same cancer type, we selected the most significant treatment association, unless a driver-treatment annotation was clearly reported in the literature.

The full catalogue of TEDs and their mutation frequencies can be found in Supplementary Table 8.

Coding mutation drivers

We used dNdScv (v0.0.1)[60] with default parameters to identify cancer driver genes from coding mutations. A global $q < 0.1$ was used as a threshold for significance. Mutation frequencies for each driver gene were extracted from the dNdScv output. We defined the mutation frequency as the number of samples bearing non-synonymous mutations.

Non-coding mutation drivers

We used ActiveDriverWGS[61] (v1.1.2, default parameters) to identify non-coding driver elements in five regulatory regions of the genome including 3′ untranslated regions (UTRs), 5′ UTRs, long non-coding RNAs, proximal promoters and splice sites. For each element category, we extracted the genomic coordinates from Ensembl v101. Each regulatory region was independently tested. To select for significant hits, we filtered on adjusted *P* values (FDR < 0.1) and a minimum of three mutated samples. We defined the mutation frequency as the number of mutated samples for each significantly mutated element in the treatment group.

CN variant drivers

We ran GISTIC2 (ref. [62]) (v2.0.23) on each of the 92 treatment and cancer-type groups using the following settings:

```
gistic2 -b <inputPath> -seg <inputSegmentation> -refgene hg19.UCSC.add_miR.140312.refgene.mat -genegistic 1 -gcm extreme -maxseg 4000 -broad 1 -brlen 0.98 -conf 0.95 -rx 0 -cap 3 -saveseg 0 -armpeel 1 -smallmem 0 -res 0.01 -ta 0.1 -td 0.1 -savedata 0 -savegene 1 -qvt 0.1.
```

The focal GISTIC peaks ($q \leq 0.1$ and <1 Mb) were then annotated with functional elements using the coordinates from Ensembl v101. The frequency differences between treated and untreated cohorts on every gene was assessed with Fisher's exact test as described above. For this, we first calculated the focal amplification and deep depletion status of every gene within each sample. A gene was amplified when the ploidy level of the gene was 2.5 ploidy levels higher than its genome-wide mean ploidy level (as measured by PURPLE), and deleted when the gene ploidy level was lower than 0.3 (that is, deep deletion). We observed that the majority of the peaks contained multiple significant gene candidates (after multiple correction $q < 0.05$) and therefore we retained the gene most closely positioned to the peak summit, which is the most significantly enriched region across the treated samples. Next, we also found recurrent peaks across multiple treatment groups per cancer type that are not, or less, present in the untreated control

group because most of the Hartwig samples have received multiple treatment types. We therefore merged peaks with overlapping ranges to produce a single peak per genomic region per cancer type. For each collapsed peak, we selected the treatment type showing the lowest *q* value for the gene near the peak summit. Deletion and amplification peaks were processed separately.

Group-level aggregation of treatment resistance-associated variants

To estimate the contribution of TEDs to the total number of drivers per sample in the metastatic cohort, we excluded any TED from the catalogue of driver mutations (see the above section 'Driver alterations') in a cancer-type-specific, gene-specific and driver-type-specific manner.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Metastatic WGS data and metadata from the Hartwig Medical Foundation are freely available for academic use through standardized procedures. Request forms can be found at https://www.hartwigmedicalfoundation.nl/en/data/data-acces-request/. Somatic variant calls, gene driver lists, CN profiles and other core data of the PCAWG cohort generated by the Hartwig analytical pipeline are available for download at https://dcc.icgc.org/releases/PCAWG/Hartwig. Researchers will need to apply to the ICGC data access compliance office (https://daco.icgc-argo.org) for the ICGC portion of the dataset. Similarly, users with authorized access can download the TCGA portion of the PCAWG dataset at https://icgc.bionimbus.org/files/5310a3ac-0344-458a-88ce-d55445540120. Additional information on accessing the data, including raw read files, can be found at https://docs.icgc.org/pcawg/data/. References and download links to the original independent datasets used in

the analyses are included in each of the pertinent sections of the Methods and Supplementary Notes, and a full list of all datasets used in the present study can be found in the data availability section of the Reporting Summary file.

Code availability

The Hartwig analytical processing pipeline is available (https://github.com/hartwigmedical/pipeline5) and implemented in Platinum (https://github.com/hartwigmedical/platinum). The source code to reproduce the analysis of the manuscript is available in the following repository: https://github.com/UMCUGenetics/primary-met-wgs-comparison.

References

1. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168**, 670–691 (2017).

2. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. *Nature* **529**, 298–306 (2016).

3. Welch, D. R. & Hurst, D. R. Defining the hallmarks of metastasis. *Cancer Res.* **79**, 3011–3027 (2019).

4. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

5. Priestley, P. et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).

6. Alexander, S. & Friedl, P. Cancer invasion and resistance: interconnected processes of disease progression and therapy failure. *Trends Mol. Med.* **18**, 13–26 (2012).

7. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169–175 (2016).

8. Massagué, J., Batlle, E. & Gomis, R. R. Understanding the molecular mechanisms driving metastasis. *Mol. Oncol.* **11**, 3–4 (2017).

9. Birkbak, N. J. & McGranahan, N. Cancer genome evolutionary trajectories in metastasis. *Cancer Cell* **37**, 8–19 (2020).

10. Weiss, F., Lauffenburger, D. & Friedl, P. Towards targeting of shared mechanisms of cancer metastasis and therapy resistance. *Nat. Rev. Cancer* **22**, 157–173 (2022).

11. Bertucci, F. et al. Genomic characterization of metastatic breast cancers. *Nature* **569**, 560–564 (2019).

12. Robinson, D. R. et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).

13. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).

14. Cameron, D. L. et al. GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol.* **22**, 202 (2021).

15. Garcia-Prieto, C. A., Martínez-Jiménez, F., Valencia, A. & Porta-Pardo, E. Detection of oncogenic and clinically actionable mutations in cancer genomes critically depends on variant calling tools. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btac306 (2022).

16. Nguyen, B. et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.e11 (2022).

17. Roepman, P. et al. Clinical validation of whole genome sequencing for cancer diagnostics. *J. Mol. Diagn.* **23**, 816–833 (2021).

18. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).

19. Steiner, T. et al. Gain in chromosome 8q correlates with early progression in hormonal treated prostate cancer. *Eur. Urol.* **41**, 167–171 (2002).

20. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).

21. Steele, C. D. et al. Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).

22. Gemble, S. et al. Genetic instability from a single S phase after whole-genome duplication. *Nature* **604**, 146–151 (2022).

23. Donehower, L. A. et al. Integrated analysis of TP53 gene and pathway alterations in The Cancer Genome Atlas. *Cell Rep.* **28**, 1370–1384.e5 (2019).

24. Kocakavuk, E. et al. Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat. Genet.* **53**, 1088–1096 (2021).

25. Christensen, S. et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat. Commun.* **10**, 4571 (2019).

26. Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836.e16 (2019).

27. Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).

28. Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).

29. Abascal, F. et al. Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).

30. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

31. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat. Rev. Cancer* **9**, 302–312 (2009).

32. Berges, R. R. et al. Implication of cell kinetic changes during the progression of human prostatic cancer. *Clin. Cancer Res.* **1**, 473–480 (1995).

33. Nguyen, L., W M Martens, J., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**, 5584 (2020).

34. Menghi, F. et al. The tandem duplicator phenotype is a prevalent genome-wide cancer configuration driven by distinct gene mutations. *Cancer Cell* **34**, 197–210.e5 (2018).

35. Degasperi, A. et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat. Cancer* **1**, 249–263 (2020).

36. Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).

37. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat. Rev. Clin. Oncol.* **12**, 573–583 (2015).

38. Watson, P. A., Arora, V. K. & Sawyers, C. L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).

39. Yun, C.-H. et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl Acad. Sci. USA* **105**, 2070–2075 (2008).

40. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).

41. Watson, R. G. et al. Amplification of thymidylate synthetase in metastatic colorectal cancer patients pretreated with 5-fluorouracil-based chemotherapy. *Eur. J. Cancer* **46**, 3358–3364 (2010).

42. Yang, L. et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598–602 (2013).

43. Xiao, Y., Lin, F.-T. & Lin, W.-C. ACTL6A promotes repair of cisplatin-induced DNA damage, a new mechanism of platinum resistance in cancer. *Proc. Natl Acad. Sci. USA* **118**, e2015808118 (2021).

44. Formisano, L. et al. Aberrant FGFR signaling mediates resistance to CDK4/6 inhibitors in ER^+ breast cancer. *Nat. Commun.* **10**, 1373 (2019).

45. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

46. GLASS Consortium. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro Oncol.* **20**, 873–884 (2018).

47. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).

48. Moldoveanu, D. et al. Spatially mapping the immune landscape of melanoma using imaging mass cytometry. *Sci. Immunol.* **7**, eabi5072 (2022).

49. Martínez-Jiménez, F. et al. Genetic immune escape landscape in primary and metastatic cancer. *Nat. Genet.* https://doi.org/10.1038/s41588-023-01367-1 (2023).

50. Nguyen, L., Van Hoeck, A. & Cuppen, E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. *Nat. Commun.* **13**, 4013 (2022).

51. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

52. Signal. *Signal* https://signal.mutationalsignatures.com/explore/study/6?mutationType=1 (accessed 1 February 2023).

53. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).

54. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

55. Shale, C. et al. Unscrambling cancer genomes via integrated analysis of structural variation and copy number. *Cell Genomics* **2**, 100112 (2022).

56. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

57. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* **2017**, PO.17.00011 (2017).

58. Griffith, M. et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **49**, 170–174 (2017).

59. Tamborero, D. et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, 25 (2018).

60. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).

61. Zhu, H. et al. Candidate cancer driver mutations in distal regulatory elements and long-range chromatin interaction networks. *Mol. Cell* **77**, 1307–1321.e10 (2020).

62. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).

Acknowledgements

This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation and the Center of Personalized Cancer Treatment (CPCT) have made available to the study. We thank R. Janssen for the technical assistance in the collection and processing of the PCAWG raw sequencing data using the Hartwig tumour analytical pipeline; USEQ from UMC Utrecht for providing high-performance computing power; J. Mateo for his valuable scientific input; L. Stein and L. Xiang for their assistance in the publication of the reprocessed ICGC part of the PCAWG dataset; R. Grossman, C. Meyer and T. Simmons for their assistance in the publication of the reprocessed TCGA part of the PCAWG; D. Quigley for providing the somatic mutation calls and age at biopsy of the patients with prostate cancer from their study; S. Turajlic and H. Pallikonda for their responsiveness concerning the TRACERx renal dataset; and P. Wolfe and other staff of the Hartwig Medical Foundation team for aligning the processing of PCAWG and Hartwig datasets.

Author information

Author notes

1. These authors contributed equally: Ali Movasati, Sascha Remy Brunner, Luan Nguyen

Authors and Affiliations

1. Center for Molecular Medicine, Oncode Institute, University Medical Center Utrecht, Utrecht, The Netherlands

   Francisco Martínez-Jiménez, Ali Movasati, Sascha Remy Brunner, Luan Nguyen, Edwin Cuppen & Arne Van Hoeck

2. Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

   Francisco Martínez-Jiménez

3. Hartwig Medical Foundation, Amsterdam, The Netherlands

   Francisco Martínez-Jiménez & Edwin Cuppen

4. Hartwig Medical Foundation Australia, Sydney, New South Wales, Australia

   Luan Nguyen & Peter Priestley

Contributions

F.M.-J., A.V.H. and E.C. conceptualized the study. F.M.-J., A.M., S.R.B., L.N., P.P. and A.V.H. came up with the methodology. F.M.-J., A.M., S.R.B., L.N., P.P. and A.V.H. provided software. F.M.-J., A.M., S.R.B., L.N., P.P. and A.V.H. validated results. F.M.-J., A.M., S.R.B., L.N., P.P., A.V.H. and E.C. conducted formal analysis. F.M.-J., A.M., S.R.B., L.N., P.P. and A.V.H. performed the investigations and provided resources. F.M.-J., A.M., S.R.B., L.N., P.P. and A.V.H. curated the data. F.M.-J. and A.V.H. wrote the original draft of the manuscript. F.M.-J., A.M., S.R.B., L.N., P.P., A.V.H. and E.C. reviewed and edited the manuscript. F.M.-J., A.M., S.R.B., L.N., A.V.H. and E.C. performed visualization. F.M.-J., A.V.H. and E.C. supervised the project. A.V.H. and E.C. provided project administration. E.C. acquired funding.

Corresponding author

Correspondence to Edwin Cuppen.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Peter Van Loo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Cohort overview and global genomic features.

a) Workflow of the unified processing pipeline used in this study for Hartwig (left) and PCAWG (right) WGS samples. First, PCAWG tumor and matched normal raw sequencing files were gathered and re-processed using the Hartwig tumor analytical pipeline. Next, the output of tumor samples that were correctly processed by the pipeline were further subjected to a strict quality control filtering. As a result, a total of 7,108 samples from 71 cancer types compose the harmonized dataset. 5,365 patient tumor samples from 23 cancer types with sufficient representation in both primary and metastatic datasets were selected for this study. **b)** Tumor clonality according to the metastatic biopsy location in kidney renal clear cell carcinoma, lung adenocarcinoma, prostate carcinoma and skin melanoma. N, number of samples in the group. p, two-sided Mann-Whitney p-value. Box-plots: center line, median; box limits, first and third quartiles; whiskers, lowest/highest data points at first quartile minus/plus 1.5× IQR. **c)** Left, similar to Fig. 1d only including non-WGD tumors. Right, similar to Fig. 1e for non-WGD tumors. **d)** Equivalent to c), but limited to WGD tumors. '*', two-sided Mann-Whitney adjusted p-value < 0.01 for continuous variables and two-sided Fisher's exact test adjusted p-value < 0.01 for *TP53*.

Extended Data Fig. 2 Mutation burden and mutational signatures.

a) double-base substitutions (DBSs, top) and indels (IDs, bottom) mutational spectra of metastatic and primary tumors. Patients are ordered according to their TMB burden. **b)** Moon plot representing the DBS burden differences attributed to each mutational signature in metastatic (left) and primary (right) tumors. Edge thickness and colors represent significant differences (Mann-Whitney adjusted p-value<0.05, ±1.4x fold change) and the direction of the enrichment, respectively. The size of circles are proportionate to the mutation burden difference. Right bars, number of metastatic cancer types with a mutational signature significant enrichment. Top stacked bars represent the cumulative signature exposure difference. Thicker bar edge lines represent significance. Bars are coloured according to the annotated etiology. Only mutational signatures with known etiology or with at least one cancer type with significant metastatic enrichment are included. **c)** analogous representation for IDs. **d)** Volcano plot representing the mutational signature hypermutation (>10,000 mutations for SBS, >500 for DBS, and >1000 for ID) prevalence comparison between primary and metastatic tumor patients. Y-axis, $\log_{10}$(two-sided Mann-Whitney adjusted p-value). X-axis, effect size as Cramer's V. Each dot represents a mutational signature in a cancer type. Dots are coloured according to the mutation type. Diff., difference. Muts. mutations. Sig., mutational signature. Mut. mutational. Susp., suspected. Def., deficiency.

Extended Data Fig. 3 Mutational signature relative contribution comparison.

a) From top to bottom, moon plot representing the SBS, DBS and ID relative contribution differences attributed to each mutational signature. The size of circles are proportional to the relative mutation burden difference. Top stacked bars represent the relative signature exposure difference. Thicker bar edge lines represent significance (two-sided Mann-Whitney adjusted p-value < 0.05 and ≥1% difference in relative contribution). Bars are coloured according to the annotated etiology. Right bars, number of metastatic cancer types with a mutational signature significant enrichment. Only mutational signatures with known etiology or with at least one cancer

type with significant enrichment are included. Diff., difference. Muts. mutations. Sig., mutational signature. Mut. mutational. Susp., suspected. Def., deficiency.

Extended Data Fig. 4 Age-corrected SBS1 mutation burden in primary and metastatic tumors.

a) Linear regression of the SBS1 mutation burden (y-axis) and patient's age at biopsy (x-axis) in primary and metastatic cancer across the 23 cancer types. The median trendline and 99% confidence intervals of the linear regression are represented as a solid line and the adjacent shaded area, respectively. The mean fold change, mean SBS1 increase per year and one sided Mann-Whitney p-value are only displayed in cancer types with an age-independent significantly different primary and metastatic distribution. Red labels, significant increase in metastatic tumors. Blue, control cancer types. R_{met} and R_{prim}, Pearson correlation coefficient of the metastatic and primary linear regressions, respectively. **b)** Analogous representation for independent linear regressions for breast cancer subtypes. **c)** Relative to a) for ploidy corrected SBS1 in the tumor types of interest. **d)** Relative to a) for ploidy corrected SBS5/40 counts in the tumor types of interest. **e)** Depiction illustrating the potential effect of an increased cell division rate in metastatic tumors compared to primary and its expected impact on the SBS1 variant allele frequency (VAF) distribution. Partially created BioRender.com. **f)** Comparison of global SBS1 clonality ratios between primary and metastatic in breast, prostate, kidney renal clear cell, thyroid, colorectal and ovarian serous carcinomas. Boxplots are defined as in Fig. 1. P, two-sided Mann-Whitney p-value. N, number of samples. **g)** Spearman correlation analysis of the mean SBS1 year burden of primary tumors (y-axis) and the mean metastatic SBS1 fold change (x-axis) across the 15 cancer types with linear association between age and SBS1 accumulation. Vertical error bars represent the 25th and 75th percentile, respectively. Horizontal error bars represent the standard deviation of the mean fold change (metastatic divided by primary) of the SBS1 yearly mutation burden. The median trendline and 99% confidence intervals of the linear regression are represented as a solid line and the adjacent shaded area, respectively. Cancer types with a significantly different SBS1 mutation rate are marked by thicker marker borders and with red labels. Blue labels represent the control cancer types.

h) Similar but using SBS1 year mutation rate from ref. [30]. To derive vertical and horizontal error bars in panels g) and h) all tumor samples from the primary and metastatic cohorts from panel a) (see Methods for inclusion criteria) were included in the analysis. The number of included samples per cancer type and cohort are available in Supplementary Table 4. Muts, mutations.

Extended Data Fig. 5 Structural variant burden.

a) Top rectangles represent the four genomic instability features defined in Fig. 1e. A red background represents significant enrichment in the metastatic cohort (two-sided Mann-Whitney adjusted p-value <0.01). S-plots, cumulative distribution function plot (samples ranked independently for each SV type) of tumor mutation burden for each cancer type for (from top to the bottom) the aggregated structural variant (SV) burden, small deletions (<10kb), large deletions (>=10kb), small duplications (<10kb), large duplications (>=10kb), complex events (<20 breakpoints), complex events (>=20 breakpoints) and LINEs insertions. Horizontal lines represent median values. Backgrounds are coloured according to the relative enrichment, defined as: $\log_{10}$(median SV type burden in metastatic tumors + 1) − $\log_{10}$(median SV type burden in primary tumors + 1). Fold change labels and coloured backgrounds are displayed when Mann-Whitney comparison renders a significant q-value < 0.05. Fold change labels are displayed with '>' when the SV burden for primary tumors is 0 (see Methods for more details). For each cancer type, bottom bar plots represent the relative fraction of each SV type in the metastatic (left) and primary (right) datasets. **b)** SV length frequency distribution of deletions (left panel) and duplications (middle panel). Right panel shows the frequency distribution of the number of linked breakpoints for complex SVs. Dashed vertical lines represent the chosen threshold to separate between short and large deletions, duplications and complex SVs, respectively.

Extended Data Fig. 6 Structural variant burden associated genomic features.

a) Volcano plot representing the cancer type-specific regression coefficients (x-axis) and significance (y-axis, measured by the linear regression model

coefficient p-value) of clinical and genomic features against the number of small deletions. Each dot represents one feature in one cancer type. Labels are coloured according to the feature category. Dots are coloured by the frequency enrichment in metastatic (purple) or primary (orange) patients. Analogous panels are displayed for **b)** large deletions, **c)** short duplications, **d)** large duplications, **e)** short complex SVs, **f)** large complex SVs and **g)** LINEs. **h)** Lollipop plots representing the regression coefficients (left, relative to panel b. x-axis) and metastatic enrichment (right, relative to dots colour from panel b.) of features associated with small deletions. Only significant features (LM>0.0, LM coefficient p-value < 0.01 and with independent significance in primary or metastatic tumors) enriched in metastatic tumor patients (enrichment > 0.0) are displayed. **i), j)** and **k)** are identical but referring to large deletions, small duplications and large duplications, respectively. LM, linear model. Coef, coefficient.

Extended Data Fig. 7 Driver landscape and drivers per patient.

a) Cancer type-specific distribution of number of driver alterations, amplifications, deletions and mutations per patient in primary (top) and metastatic (bottom). Black dots represent the mean values. Labels display mean differences (metastatic - primary) in cancer types with a significant difference. "*", two-sided Mann-Whitney adjusted p-value < 0.01. **b)** Volcano plots representing the cancer type-specific enrichment (x-axis) and significance (y-axis, FDR adjusted two-tailed Fisher's exact test p-value) of driver genes between primary and metastatic cohorts. From left to right, amplification drivers, biallelically deleted drivers and mutated driver genes. BRCA, Breast carcinoma. KIRC, kidney renal clear cell carcinoma. OV, Ovarian serous adenocarcinoma. PRAD, Prostate carcinoma. SKCM, Skin melanoma. THCA, Thyroid carcinoma. LIHC, Hepatocellular carcinoma. PANET, pancreatic neuroendocrine.

Extended Data Fig. 8 Therapeutic actionability of variants.

a) Cancer type-specific fraction of primary (top) and metastatic (bottom) patients with reported therapeutically actionable variants. For each patient the variant with the greatest level of evidence was considered. Bars are coloured according to the variant actionability tiers. Fold change (i.e.,

metastatic divided by primary fraction) labels are displayed in cancer types with a significant proportional increase (two-sided Fisher's exact test adjusted p-value < 0.05). Purple edgelines highlight significant increase in metastatic A-on label fraction patients. **b)** Primary (left) and metastatic (right) alteration frequency of actionable variants with a high discrepancy (>5% frequency difference) from cancer types with a global significant increase of actionable variants in metastatic tumor patients from panel a). "*", a two-sided Fisher's exact test adjusted p-value < 0.05. Text boxes include the associated treatments for alterations with a significant mutation frequency increase in metastatic tumor patients.

Extended Data Fig. 9 Treatment enriched drivers.

a) Visual depiction of the analytical framework to identify treatment enriched drivers (TEDs). Example, identification of TEDs in the 354 breast carcinoma patients treated with aromatase inhibitors. (1), identification of cancer driver genes from coding mutations (green), non-coding mutations (soft green), copy number amplifications (red) and deletions (blue). (2), for each driver gene, comparison of the alteration frequency in treated and untreated patients. (3) and (4), annotation of TEDs with type of enrichment and orthogonal evidence **b)** Side by side alteration frequency comparison between treated (right bar) and untreated (left bar) patients for all treatment-exclusive and **c)** treatment-enriched TEDs. **d)** Distribution of mutations along the *AR* protein sequence in prostate cancer patients treated with androgen deprivation (top) and untreated (bottom). Pfam domains are represented as rectangles. Mutations are coloured according to the consequence type. **e)** Distribution of focal copy number gains in chromosome X in prostate untreated patients (bottom) and treated with androgen deprivation (top). *AR* coding region and the promoter region are highlighted. **f)** Distribution of mutations along the *ESR1* protein sequence in breast carcinoma patients treated with aromatase inhibitors (top) and untreated (bottom). **g)** Distribution of mutations along the *EGFR* protein in lung adenocarcinoma patients treated with EGFR inhibitors (top) and untreated (bottom). Pfam domains are represented as rectangles. **h)** Distribution of focal copy number gains in chromosome 7 in lung adenocarcinoma untreated patients (bottom) and treated with anti-EGFR (top). *EGFR*, *MET* and *CDK6* genomic locations are highlighted. **i)**

Distribution of focal copy number gains in chromosome chr18p:1Mb-8Mb in breast carcinoma untreated patients (bottom) and treated with pyrimidine antagonists (top). TYMS genomic location is highlighted. **j)** Similar to f) but representing ultra-focal (shorter than 3Mbs) MYC and PRNCR1 amplifications in chromosome 8q. In all copy number gain plots each bin represents 100Kbs. Mb, megabase. Kb, kilobase.

Extended Data Fig. 10 Pan-cancer differences between primary and metastatic tumors.

a) Stacked plot representing the qualitative differences of the eight studied genomic features across the 23 cancer types included in this study. Cancer types are sorted in ascending order according to the cumulative number of diverging genomic features between primary and metastatic tumors. Each horizontal track represents a genomic feature. The presence (and height) of each feature for a specific cancer type correlates with the magnitude of the observed differences.

Supplementary information

Supplementary Information

This file contains Supplementary Notes 1-4

Reporting Summary

Peer Review File

Supplementary Data 1

Mutation profiles of de novo signatures that were extracted with non-negative matrix factorization.

Supplementary Table 1

Primary and metastatic cohorts metadata.

Supplementary Table 2

Karyotype and genomic instability measurements.

Supplementary Table 3

Mutational signatures absolute and relative contribution.

Supplementary Table 4

SBS1 mutation rate.

Supplementary Table 5

Structural variants.

Supplementary Table 6

Driver alterations.

Supplementary Table 7

Therapeutic actionability of variants.

Supplementary Table 8

Treatment associated drivers (TEDs).

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's

Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Martínez-Jiménez, F., Movasati, A., Brunner, S.R. *et al.* Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* **618**, 333–341 (2023). https://doi.org/10.1038/s41586-023-06054-z

- Received: 28 June 2022
- Accepted: 05 April 2023
- Published: 10 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06054-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **[Genetic immune escape landscape in primary and metastatic cancer]**

    - Francisco Martínez-Jiménez
    - Peter Priestley
    - Edwin Cuppen

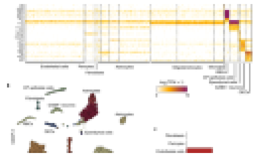
    Nature Genetics (2023)

Genetic immune escape landscape in primary and metastatic cancer

- Francisco Martínez-Jiménez
- Peter Priestley
- Edwin Cuppen

Nature Genetics Article Open Access 10 May 2023

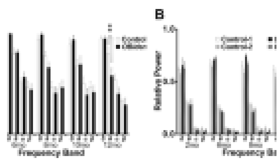
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06054-z



A shared disease-associated oligodendrocyte signature among multiple CNS pathologies

27 June 2022

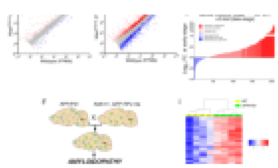
Mor Kenigsbuch, Pierre Bost, … Ido Amit



Corticohippocampal Dysfunction In The OBiden Mouse Model Of Primary Oligodendrogliopathy

31 October 2018

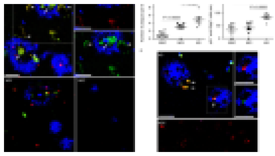
Daniel Z. Radecki, Elizabeth L. Johnson, … Alexander Gow



Reactive astrocytes acquire neuroprotective as well as deleterious signatures in response to Tau and Aß pathology

10 January 2022

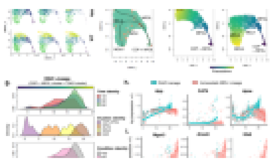
Zoeb Jiwaji, Sachin S. Tiwari, … Giles E. Hardingham



[Senolytic therapy alleviates Aβ-associated oligodendrocyte progenitor cell senescence and cognitive deficits in an Alzheimer's disease model]

01 April 2019

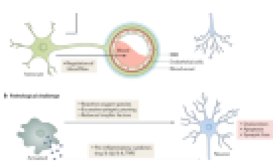
Peisu Zhang, Yuki Kishimoto, … Mark P. Mattson



[Single-cell RNA-sequencing identifies disease-associated oligodendrocytes in male APP NL-G-F and 5XFAD mice]

13 February 2023

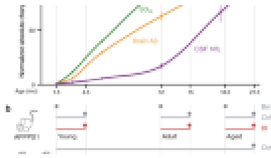
Hanseul Park, Byounggook Cho, … Jongpil Kim



[Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here?]

14 December 2020

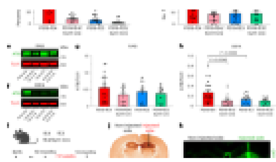
Fangda Leng & Paul Edison



[Experimental evidence for temporal uncoupling of brain Aβ deposition and neurodegenerative sequelae]

28 November 2022

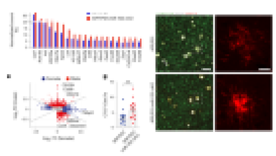
Christine Rother, Ruth E. Uhlmann, … Mathias Jucker



[Neuronal APOE4 removal protects against tau-mediated gliosis, neurodegeneration and myelin deficits]

20 February 2023

Nicole Koutsodendris, Jessica Blumenfeld, … Yadong Huang



[Identification of a protective microglial state mediated by miR-155 and interferon-γ signaling in a mouse model of Alzheimer's disease]

08 June 2023

Zhuoran Yin, Shawn Herron, … Oleg Butovsky

- Article
- Open Access

- Published: 31 May 2023

Myelin dysfunction drives amyloid-β deposition in models of Alzheimer's disease

- Constanze Depp ,
- Ting Sun,
- Andrew Octavian Sasmita,
- Lena Spieth,
- Stefan A. Berghoff,
- Taisiia Nazarenko,
- Katharina Overhoff,
- Agnes A. Steixner-Kumar,
- Swati Subramanian,
- Sahab Arinrad,
- Torben Ruhwedel,
- Wiebke Möbius,
- Sandra Göbbels,
- Gesine Saher,
- Hauke B. Werner,
- Alkmini Damkou,
- Silvia Zampar,
- Oliver Wirths,
- Maik Thalmann,
- Mikael Simons,
- Takashi Saito,
- Takaomi Saido,
- Dilja Krueger-Burg,
- Riki Kawaguchi,
- Michael Willem,
- Christian Haass,
- Daniel Geschwind,
- Hannelore Ehrenreich,

- Ruth Stassart &
- …
- Klaus-Armin Nave

Show authors

Nature **volume 618**, pages 349–357 (2023)

- 14k Accesses
- 298 Altmetric
- Metrics details

Abstract

The incidence of Alzheimer's disease (AD), the leading cause of dementia, increases rapidly with age, but why age constitutes the main risk factor is still poorly understood. Brain ageing affects oligodendrocytes and the structural integrity of myelin sheaths[1], the latter of which is associated with secondary neuroinflammation[2,3]. As oligodendrocytes support axonal energy metabolism and neuronal health[4,5,6,7], we hypothesized that loss of myelin integrity could be an upstream risk factor for neuronal amyloid-β (Aβ) deposition, the central neuropathological hallmark of AD. Here we identify genetic pathways of myelin dysfunction and demyelinating injuries as potent drivers of amyloid deposition in mouse models of AD. Mechanistically, myelin dysfunction causes the accumulation of the Aβ-producing machinery within axonal swellings and increases the cleavage of cortical amyloid precursor protein. Suprisingly, AD mice with dysfunctional myelin lack plaque-corralling microglia despite an overall increase in their numbers. Bulk and single-cell transcriptomics of AD mouse models with myelin defects show that there is a concomitant induction of highly similar but distinct disease-associated microglia signatures specific to myelin damage and amyloid plaques, respectively. Despite successful induction, amyloid disease-associated microglia (DAM) that usually clear amyloid plaques are apparently distracted to nearby myelin damage. Our data suggest a working model whereby age-dependent structural defects of myelin promote Aβ

plaque formation directly and indirectly and are therefore an upstream AD risk factor. Improving oligodendrocyte health and myelin integrity could be a promising target to delay development and slow progression of AD.

Main

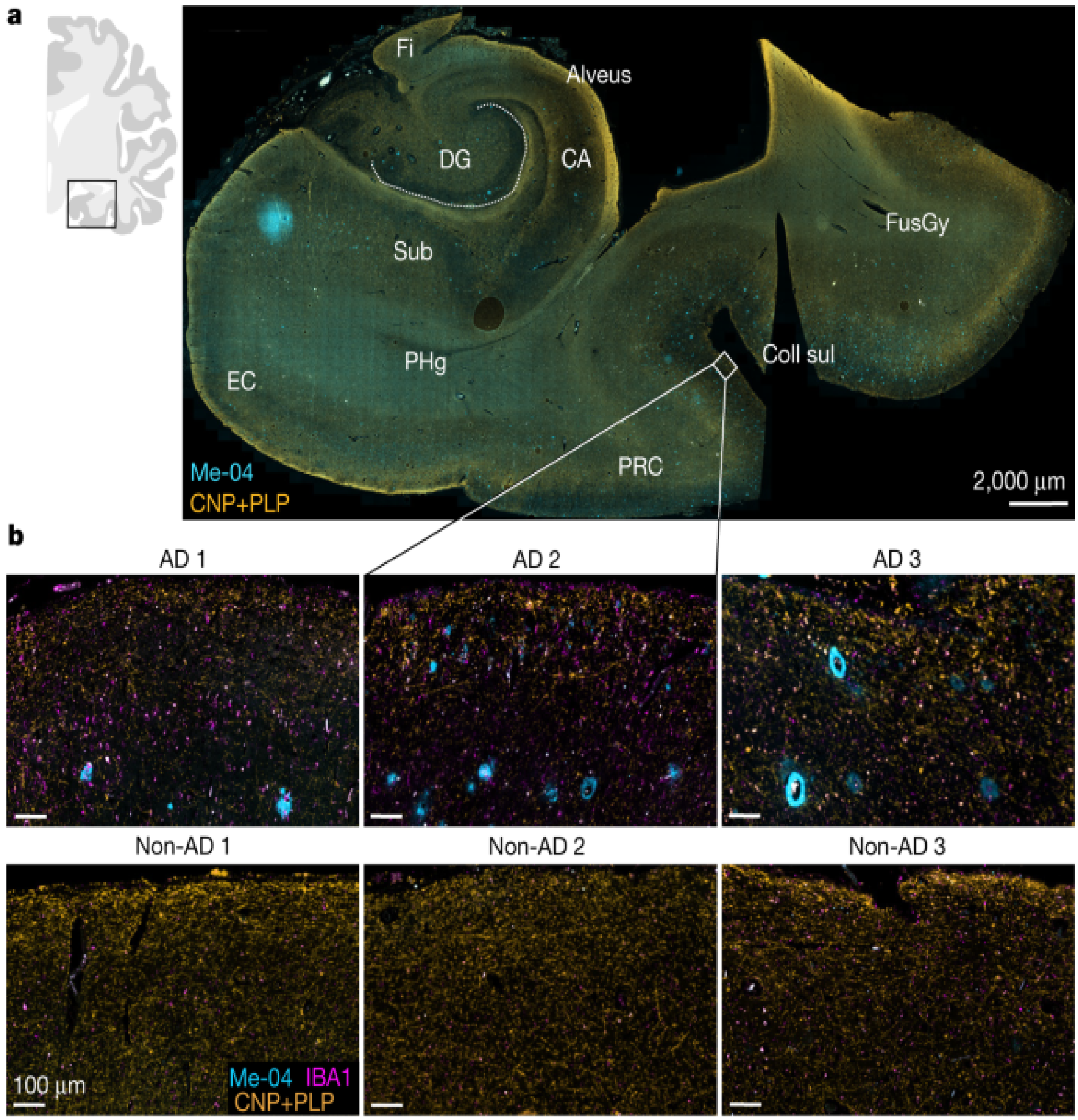
The pathology of AD is characterized by the deposition of Aβ plaques and neurofibrillary tangles primarily in the neocortex and hippocampus. According to the amyloid hypothesis of AD, the build-up of Aβ initiates a cascade of harmful events that lead to neuronal dysfunction. More than lifestyle choices and genetic predisposition, old age is the primary risk factor for AD development, but exactly how brain ageing is linked to amyloid deposition is unclear. Myelin, a spirally wrapped and compacted glial membrane, enhances axonal conduction speed, and its non-compacted regions allow oligodendrocytes to provide metabolic support to the neuronal compartment[7]. The specific cellular architecture of myelin makes protein and lipid turnover challenging and slow[8,9]. This, together with the long lifetime of oligodendrocytes[10], might explain the structural deterioration of myelin with age[1]. We speculated that the breakdown of myelin integrity in the ageing brain acts as a driving force for Aβ deposition. We tested this hypothesis in mouse models of amyloidosis, in which we used genetic and pharmacological manipulation to introduce various degrees of myelin dysfunction.

Myelin decline in patients with AD

Macroscopic-scale brain imaging studies have suggested that cortical myelin damage occurs in the preclinical phase of AD[11,12,13,14]. Microscopy-based evidence of this damage, however, remains scarce. We therefore used immunofluorescence to analyse intracortical myelin integrity in a small group of patients with AD, focusing on the trans-entorhinal area (Fig. 1 and Extended Data Fig. 1a,b). We observed a decline in intracortical myelin density in autopsy samples from patients with AD that were not limited to the immediate vicinity of plaques. Sites of myelin loss were also associated with increased numbers of $IBA1^{+}$ microglia (Extended Data Fig. 1a,b). Correlations in human neuropathology cannot determine whether myelin

loss is the cause or consequence of neuronal AD pathology such as axonal decay. To clarify this issue, we used two different experimental animal models of AD (5×FAD and APP^{NLGF}; Extended Data Fig. 1c) to investigate whether loss of myelin can act as an upstream driver of amyloidosis. Specifically, we combined the AD mouse models with mice that develop subtle myelin disintegration that is induced using genetic methods (Extended Data Fig. 1c). Mice with loss of function (null mutants) of the myelin architectural proteins CNP and PLP display minor structural myelin defects: lack of CNP causes the collapse of cytosolic channels within the myelin sheath[15] associated with axonal swellings[16]. PLP-deficient myelin lacks physical stability, impairs axonal energy metabolism[17] and causes axonopathy at higher age[4]. We first compared normal brain ageing of C57BL/6 wild-type (WT) mice with that of $Cnp^{-/-}$ mice and $Plp^{-/y}$ mice on the same background. Immunostaining showed that $Cnp^{-/-}$ and $Plp^{-/y}$ mice have progressive gliosis in grey and white matter between the age of 3 and 6 months. The progressive gliosis was qualitatively similar to aged WT mice (24 months old) but more substantial (Extended Data Fig. 1d,e and Supplementary Fig. 2). Similar to patients with AD, the aged WT mice and more so the myelin-defective mutant mice displayed significant reductions in intracortical myelin content, especially in the upper cortical layers (Extended Data Fig. 1d,e).

Fig. 1: Myelin damage in patients with AD.

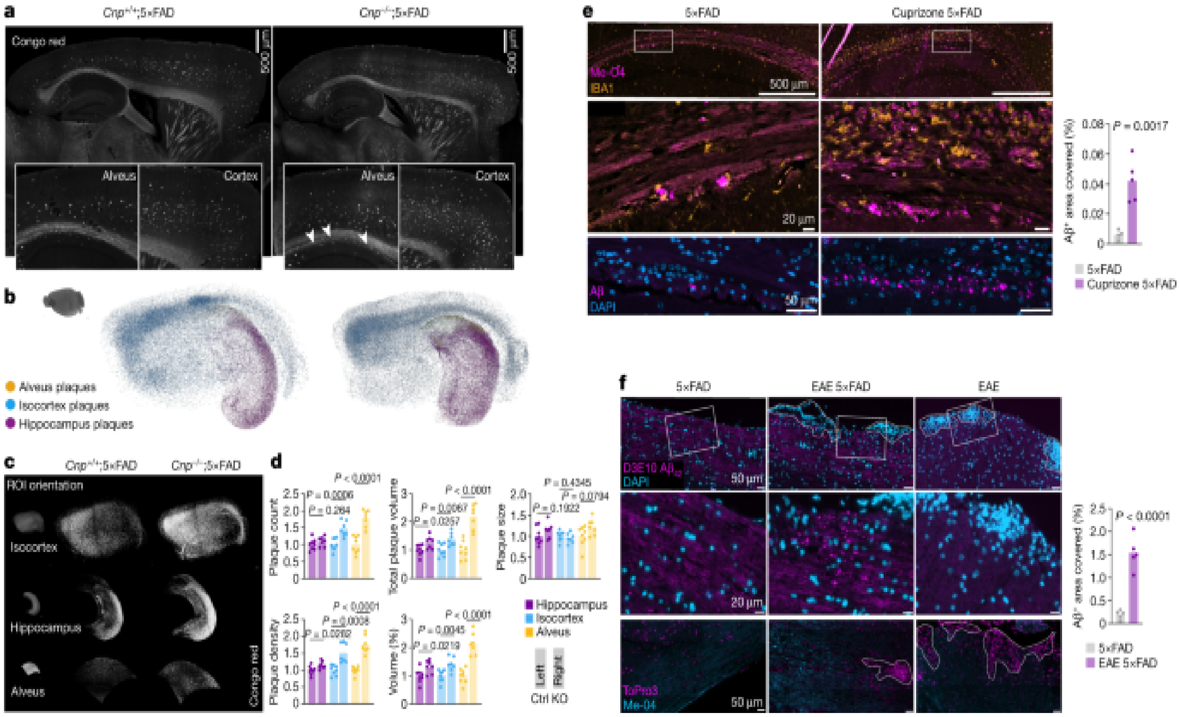


Fluorescent immunolabelling of CNP and PLP (myelinated fibres) and IBA1 (microglia) and Me-04 staining (Aβ plaques) in the medial temporal lobe of patients with AD and in unaffected individuals (non-AD). **a**, Annotated overview image of the medial temporal lobe of a patient with AD indicating the location of magnified images. **b**, Magnified images of the upper cortical layers in the transentorhinal cortex in patients with AD and in unaffected individuals ($n = 3$ per group). CA, cornu ammonis; Coll sul: collateral sulcus; DG, dentate gyrus; EC, entorhinal cortex; Fi, fimbria; FusGy, fusiform gyrus; PHg, parahippocampal gyrus; PRC, perirhinal cortex; Sub, subiculum.

Myelin defects exacerbate Aβ deposition

To determine whether such ageing-associated myelin defects can drive amyloid deposition, we crossbred *Cnp*$^{-/-}$ and *Plp*$^{-/y}$ mutants with 5×FAD mice and analysed plaque burden in the resulting offspring (Extended Data Fig. 1c). For this analysis, we optimized a Congo red in toto plaque staining and clearing protocol based on the iDISCO technique for light sheet microscopy (LSM) (Supplementary Fig. 3) to determine the amyloid plaque load in the entire brain in an unbiased fashion. When compared with 5×FAD mice, both *Cnp*$^{-/-}$;5×FAD mice and *Plp*$^{-/y}$;5×FAD mice showed marked increases in amyloid plaque load in hippocampal white matter (alveus) and cortex at 6 months of age (Fig. 2a–d and Extended Data Fig. 2a–d). In both mouse models, the effects were strongest in the alveus, a hippocampal white matter tract: here Aβ was deposited in very small aggregates, which indicated an increased formation of amyloid seeds. Increases in plaque load were observed at 6 months but not yet at 3 months of age (Extended Data Fig. 2e,f) when *Cnp*$^{-/-}$ mice and *Plp*$^{-/y}$ mice exhibited only subtle myelin defects and reactive gliosis (Extended Data Fig. 1d). Moreover, 5×FAD control mice and the myelin mutant mice did not show plaque pathology in the alveus at this age (Extended Data Fig. 2g). To exclude the possibility that these effects are specific to the 5×FAD model and *App* overexpression, we validated our findings in crossbreedings of *Cnp*$^{-/-}$ mice and the *APP*^{NLGF} knock-in model, in which three mutations in the human gene encoding amyloid precursor protein (*APP*) is expressed under the control of the endogenous *App* locus (Extended Data Fig. 2h). We also tested whether aged myelin mutant animals per se (without the human *APP* transgene) could naturally develop amyloid plaques in their lifetime. No amyloid deposits, however, were detected in 14-month old *Plp*$^{-/y}$ mice or in 22-month-old long-lived mice in which *Plp* was conditionally knocked out in the forebrain (*Emx*^{cre}*Plp*$^{fl/fl}$ mice) (Extended Data Fig. 2i,j). These results are in line with studies showing that rodent amyloid is relatively resistant to aggregation[18] and that mice overexpressing rodent Aβ do not develop amyloidosis in their lifetime[19].

Fig. 2: Dysmyelination and demyelination enhance amyloid plaque deposition in 5×FAD mice.



a–**d**, LSM analysis of 6-month-old *Cnp*$^{+/+}$;5xFAD and *Cnp*$^{-/-}$;5xFAD mouse brains stained for Congo red. **a**, Representative LSM 2D single planes. Inlays show close-up images of the cortex and alveus. Arrowheads indicate small amyloid deposits in the alveus. **b**, 3D representation of hippocampal, isocortical and alveus plaques represented as coloured centroids. **c**, 3D cropped regions of interest (ROIs) of a representative brain rendered in maximum intensity modus. **d**, 3D quantification of plaque load in the indicated ROIs normalized to the controls. $n = 8$ for control, $n = 7$ for mutant. Ctrl, control; KO, knockout. **e**, Left, 2D immunostaining images of microglia (IBA1) and Aβ plaques with Me-04 (top two rows) or antibody-labelling (bottom row) in the alveus of cuprizone-treated 5×FAD mice (cuprizone 5×FAD) and control animals (5xFAD). Right, quantification of amyloid-positive deposits in the alveus. $n = 4$ for control, $n = 5$ for cuprizone treatment. **f**, Left, 2D immunostaining images of amyloid using antibody labelling (top two rows) or the β-sheet dye Me-04 (bottom row). EAE lesions are indicated by nuclei accumulations (DAPI or ToPro3 labelling) and marked by dashed lines. EAE control animals are shown to rule out nonspecific staining of lesion sites in EAE. Right, quantification of amyloid-positive deposits in the lesion environment. $n = 5$ for control, $n = 5$ for EAE treatment. For **d**–**f**, statistical analysis: two-sided, unpaired Student's *t*-test

(*P* values are indicated in the graphs). Bars represent the means, dots represent biological replicates/mice/*n*.

Source data

Secondary myelin alterations have been described in mouse models of AD[20,21]. However, our immunostaining, western blot and electron microscopy (EM) analyses did not detect substantial myelin pathology in 6-month-old 5×FAD mice or APP^{NLGF} mice (Extended Data Fig. 3). Similarly, the subtle myelin pathology that occurs in $Cnp^{-/-}$ mice and $Plp^{-/y}$ mice was not modified by crossbreeding to the 5×FAD genotype (Extended Data Fig. 3i–l). The observed plaque-promoting effects are therefore not simply an altered secondary response of oligodendrocytes or myelin to amyloid deposits.

We also investigated whether myelin dysfunction modifies 5×FAD behavioural deficits by performing behavioural testing in the Y-maze and the elevated plus maze (Extended Data Fig. 3m–p). In both types of mazes, myelin mutant mice presented with hyperactivity that was further exacerbated in $Cnp^{-/-}$;5×FAD mice. In the elevated plus maze, $Cnp^{-/-}$5×FAD mice preferred the open arms, which is indicative of an unusual lack of anxiety. Statistical analyses confirmed supra-additive effects (interactions) of 5×FAD and especially the $Cnp^{-/-}$ genotype. Myelin defects and amyloid pathology seem to synergistically worsen behavioural deficits indicative of disinhibition, which is a neuropsychiatric symptom that is also present in patients with AD[22].

Acute demyelination drives Aβ deposition

To confirm that myelin dysfunction is an upstream driver of plaque pathology, we tested the effect of acute demyelination. Young adult 5×FAD mice were fed a cuprizone-containing diet for 4 weeks followed by a 4-week recovery period, after which the plaque load was determined by LSM. This method leads to substantial demyelination in the medial corpus callosum and the alveus accompanied by microgliosis followed by complete remyelination after 4 weeks (Extended Data Fig. 4a,b). Interpreting the outcome of this experiment was complicated, however, because the copper-chelating

properties of cuprizone interfere with plaque core formation, which depends on copper[23]. Indeed, the in toto LSM results showed that cuprizone treatment seemed to ameliorate hippocampal and cortical Aβ pathology (Extended Data Fig. 5c). However, the amyloidosis-driving effect of demyelination predominated the inhibition by copper chelation in substantially demyelinated areas, such as the hippocampal alveus. In this brain region, less-compacted Aβ plaques in cuprizone-treated 5×FAD mice were well stained with anti-Aβ antibodies, which revealed a large increase in small amyloid aggregates (Fig. 2e).

Independent evidence to support our working model was provided by results from 5×FAD mice in which experimental autoimmune encephalomyelitis (EAE) was induced. We immunized young 5×FAD animals with myelin oligodendrocyte glycoprotein (MOG) peptide and analysed their brains and spinal cords 4 weeks later (Fig. 2f and Extended Data Fig. 4d). Results from a previous study[24] of aged J20 and Tg2576 mouse models of AD showed that EAE reportedly reduced brain plaque load. However, we did not find a difference in amyloid deposition in young 5×FAD mice with EAE (Extended Data Fig. 4e) due to the lack of demyelination pathology in the brain (Extended Data Fig. 4f–h). By contrast, in the spinal cord, where demyelinating EAE pathology substantially develops (Extended Data Fig. 4i), we identified small, atypical amyloid aggregates in the peri-lesion environment, which were absent from control 5×FAD mice (Fig. 2f). We verified the presence of aggregated amyloid by staining spinal cord sections with the β-sheet dye methoxy-04 (Me-04) (Fig. 2f). Of note, in spinal cord from WT animals with EAE, no such Me-04^{+} material was found, which rules out the possibility of nonspecific detection of lipid deposits in demyelinated lesions (Fig. 2f). Spinal cord grey matter plaque deposition, as seen later in 5×FAD mice with disease progression, could not be detected in 14-week-old 5×FAD mice or in EAE 5×FAD mice (Extended Data Fig. 4i,j). Taken together, myelin defects—both chronic and acute—drive amyloid deposition in mouse models of AD, which identifies dysfunctional myelin as an upstream risk factor for amyloid deposition.

Lack of myelin ameliorates Aβ deposition

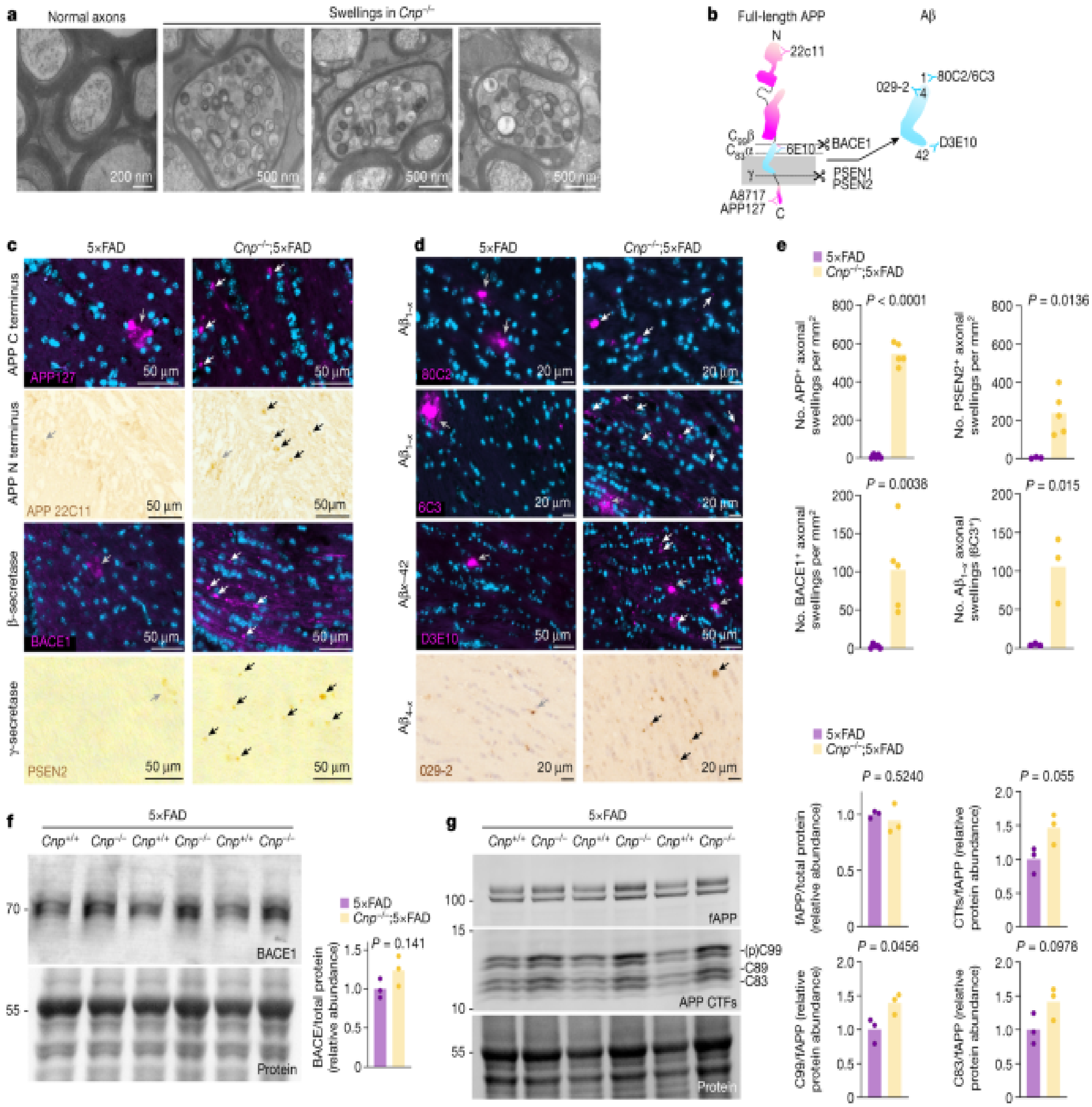
Compared to other apes, humans show a disproportional enlargement of prefrontal white matter[25]. This raises the question of whether the extent of cortical myelination per se could have a causal role in human AD. We therefore wondered what the impact of the near complete absence of cortical myelin would have on the course of amyloidosis in 5×FAD mice. To this end, we generated a line of forebrain-specific shiverer mice in which cortical axons are largely unmyelinated (*Emx*^{cre}*Mbp*$^{fl/fl}$; forebrain shiverer mice) and crossbred them to 5×FAD mice (Extended Data Fig. 5a). At 3 months of age, forebrain shiverer;5×FAD mice were strongly protected against amyloid deposition in both the hippocampus and the cortex (Extended Data Fig. 5b–d). However, at 6 months of age, this effect was largely lost, which indicated that there is a delay in plaque formation in the absence of myelin (Extended Data Fig. 5e). At 6 months of age, plaque burden in the alveus was increased in forebrain shiverer;5×FAD mice, quantitatively similar to *Cnp*$^{-/-}$ and *Plp*$^{-/y}$ crossbred mice. Qualitatively, however, alveus plaques in forebrain shiverer;5×FAD mice differed in their morphology and distribution (Extended Data Fig. 5f), which pointed towards a different mechanism of increased plaque burden in these models. We conclude that proper myelin ensheathment and healthy myelin have an initially inhibitory effect on plaque formation.

Myelin dysfunction affects APP metabolism

We next investigated how myelin defects mechanistically drive amyloidosis. In theory, the defects could either promote APP processing and Aβ generation or interfere with Aβ removal (or a combination of both). We first investigated APP metabolism in *Cnp*$^{-/-}$;5×FAD mice. Axonal swellings that stain positive for APP are prominent features of ischaemia, brain injury and myelin disorders, and we speculated that these swellings contribute to the generation of Aβ in myelin mutant animals. Stalling axonal transport can enhance amyloid production, probably by increasing BACE1 and APP interactions in axonally transported vesicles[26,27,28], and axons seem to be important sites of amyloid secretion[29,30,31]. In *Cnp*$^{-/-}$ mice and *Cnp*$^{-/-}$;5×FAD crossbred mice, numerous myelin damage-associated swellings of axonal origin were observed (Extended Data Fig. 6a–d). In 5×FAD control mice, almost no such swellings were found and axonal injury was limited to

the immediate vicinity of the plaque and arranged in a circular pattern (termed corona) (Extended Data Fig. 6a–d). High-pressure freezing EM analyses showed that myelin-damage-associated swellings were enriched in vesicular structures that were probably endosomal or lysosomal in origin (Fig. 3a), which are thought to be primary production sites of Aβ[32,33]. Using antibodies against APP processing enzymes and multiple APP- and Aβ-specific antibodies (Fig. 3b), we found that axonal swellings in the brains of *Cnp*$^{-/-}$;5×FAD mice accumulated β-secretase and γ-secretase, co-stain for APP and BACE1 and consequently stain positive for β- and γ-cleaved APP fragments and Aβ (Fig. 3c–e and Extended Data Fig. 6e–g). Immunostaining also showed increased levels of soluble APPβ in cortical tissue in *Cnp*$^{-/-}$;5×FAD mice. In this brain region, similar to white matter, numerous myelin-damage-associated swellings were observed in addition to plaque-associated swellings (Extended Data Fig. 6h). Similar observations were made in *Plp*$^{-/y}$;5×FAD mice, which were used as an independent model of myelin dysfunction (Extended Data Fig. 6i–k).

Fig. 3: Myelin dysfunction alters APP processing.



a, High-pressure freezing EM of optic nerves from 6-month-old WT and *Cnp*$^{-/-}$ mice. Nerves from *Cnp*$^{-/-}$ mice show (plaque-independent) axonal swellings with accumulation of endosomal and lysosomal structures and multivesicular bodies. The same observations were made in three independent samples per mice per group. **b**, Schematic representation of APP cleavage and binding sites of anti-APP and anti-Aβ antibodies used. **c**, Fluorescent and chromogenic immunostaining images of APP and APP cleavage enzymes BACE1 (β-secretase) and PSEN2 (as part of the γ-secretase complex) in white matter (fimbria) from *Cnp*$^{-/-}$;5×FAD mice. Grey arrowheads mark plaque-associated axonal swellings typically forming

a corona. Black and white arrowheads indicate plaque-independent axonal swellings as observed in *Cnp*$^{-/-}$ mice. **d**, Fluorescent and chromogenic immunostainings of Aβ peptides in white matter (fimbria) of *Cnp*$^{-/-}$;5×FAD mice and 5×FAD mice. Grey arrowheads indicate proper amyloid plaques, typically stained intensely. White arrows indicate swellings stained positive by the respective Aβ-antibody, but typically stained less intensely and more round in structure. **e**, Quantification of axonal swellings positive for APP, BACE1, PSEN2 and Aβ (6C3) in white matter (fimbria) of 5×FAD control mice and *Cnp*$^{-/-}$;5xFAD mice. For BACE1 and APP quantification, $n = 5$ for 5×FAD and *Cnp*$^{-/-}$;5×FAD. For PSEN2 quantification, $n = 3$ for 5×FAD and $n = 5$ for *Cnp*$^{-/-}$;5×FAD. For 6C3, $n = 3$ for 5×FAD and *Cnp*$^{-/-}$;5×FAD. **f**, Fluorescent immunoblot analysis of BACE1 levels in microdissected cortex of *Cnp*$^{-/-}$;5×FAD mice and 5×FAD mice. **g**, Fluorescent immunoblot analysis of APP fragmentation in the membrane-bound fraction of microdissected cortical tissue of *Cnp*$^{-/-}$;5×FAD mice and 5×FAD control mice. (p)C99, phospho-C99. For **f** and **g**, the molecular weight marker (in kDa) is indicated on the left. For quantifications in **f** and **g**, full-length APP (fAPP) and BACE1 levels were normalized to total protein stain. CTF levels were normalized to fAPP. $n = 3$ per group. For **e**–**g**, statistical analysis: two-sided, unpaired Student's *t*-test (*P* values are indicated in the graphs). Bars represent means, dots represent biological replicates/mice/lanes/*n*. Source data are given in Supplementary Fig. 1.

Source data

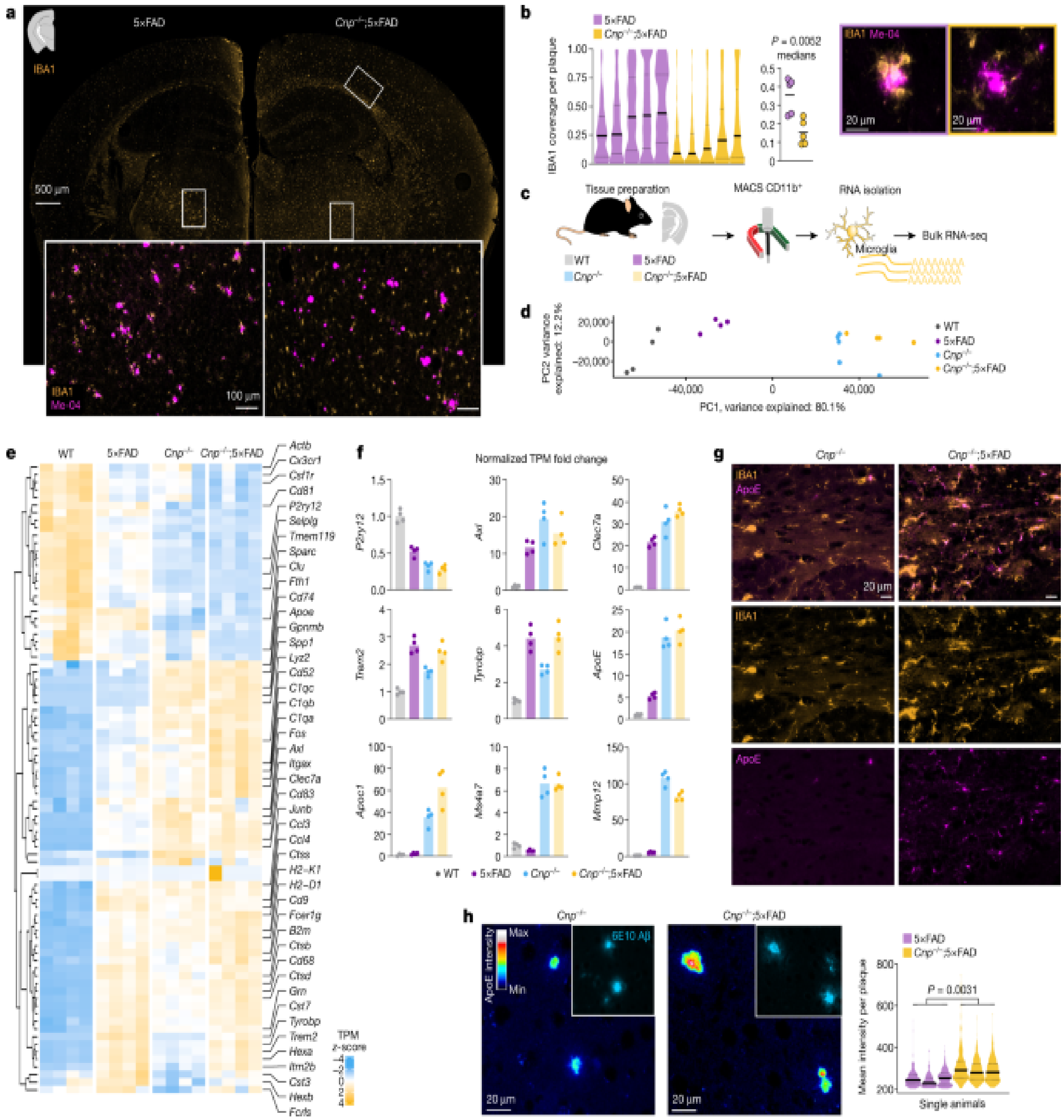
We further corroborated these findings by western blot analyses of white matter and the cortex, which showed increased levels of BACE1 (which was significant in white matter) (Fig. 3f and Extended Data Fig. 6l) in 6-month-old *Cnp*$^{-/-}$;5×FAD mice. When we investigated APP processing by immunoblotting, cortical but not white matter APP metabolism was shifted towards an increased abundance of APP carboxy-terminal fragments (CTFs), but without changes to full-length APP abundance or the α/β CTF ratio (Fig. 3g and Extended Data Fig. 6m–o). It is tempting to speculate that axonally generated CTFs in the white matter could be retrogradely transported and accumulate over time in neuronal cell bodies of the cortex, as western blotting showed that these areas are enriched in CTFs. Together,

these findings suggest that myelin defects affect APP metabolism and Aβ generation.

Myelin defects alter microglia responses

Glial cells play important parts in the clearance of myelin debris and amyloid peptides, and microglia form barriers around amyloid plaques[34]. Thus, we investigated in vitro and in vivo how these cells react to amyloid when additionally challenged with defective myelin. Prior exposure of bone-marrow-derived macrophages—a commonly used in vitro model to study AD-related phagocyte function—to myelin debris strongly inhibited amyloid phagocytosis (Extended Data Fig. 7a,b). In vivo, cortical microglia failed to cluster around amyloid plaques in *Cnp*$^{-/-}$;5×FAD mice and *Plp*$^{-/y}$5×FAD mice despite substantial increases in microglia numbers in both white and grey matter in these models (Fig. 4a–d and Extended Data Fig. 7c–k). Notably, we observed the same effect in *Cnp*$^{-/-}$;*APP*^{NLGF} crossbred mice (Extended Data Fig. 7h). An equivalent analysis of this phenotype was not possible in the primary demyelination models owing to substantial microglia and macrophage infiltration and because the primary effects in these models were not observed in cortical tissue (Extended Data Fig. 7l,m).

Fig. 4: Loss of microglial corralling around amyloid plaques in myelin mutant mice.



a, 2D immunofluorescence analysis of microglial reaction to amyloid plaques in *Cnp*$^{-/-}$;5×FAD and 5×FAD mice by IBA1 and amyloid co-staining (Me-04). **b**, Left, automated quantification of IBA1 plaque coverage in the cortex. Each violin plot represents a single animal/biological replicate/*n*. *n* = 5 per group. Black lines represent medians. In total, 2,017 individual cortical plaques were analysed in 5×FAD brain slices and 2,190 in *Cnp*$^{-/-}$;5×FAD brain slices. Middle, graph shows distribution of medians. Lines indicate means. Statistical analysis: two-sided, unpaired Student's *t*-test (*P* values are indicated on the graphs) on biological replicate data. Right,

two representative plaques for each genotype. **c**, Experimental setup for microglia bulk RNA-seq. Microglia were isolated from hemispheres of 6-month-old WT, *Cnp*$^{-/-}$, 5×FAD and *Cnp*$^{-/-}$;5×FAD animals and subjected to RNA-seq (biological replicates, $n = 4$ for each genotype). **d**, PCA was used for evaluating relative distances between normalized RNA transcripts per kilobase million (TPM) profiles. Dots represent biological replicates. Principal component 1 (PC1) explained 80.1% of data variability and strongly reflected *Cnp*$^{-/-}$-dominated microglia transcriptome changes. **e**, Heatmap of top genes contributing to PC1 variability. **f**, Normalized expression level for selected genes (homeostatic marker *P2ry12*, DAM signature *Trem2*, *Tyrobp*, *Axl* and *Clec7a*, and differential regulated genes in *Cnp*$^{-/-}$;5×FAD mice (*Apoc1*, *ApoE, Ms4a7* and *Mmp12*)). Bars represent means, dots represent mice/biological replicates/*n* ($n = 4$ per group). **g**, ApoE enrichment in microglia in white matter from *Cnp*$^{-/-}$;5×FAD mice. Immunofluorescence staining of ApoE and IBA1 to mark microglia. The same observations were made in three independent samples per mice per group. **h**, Left, microscopy analysis of plaque-associated ApoE in cortical plaques in 5×FAD mice and *Cnp*$^{-/-}$;5xFAD mice pseudo-coloured according to a rainbow lookup table. Right, violin plots of mean ApoE fluorescence intensity per plaque. Each violin plot represents a single animal. Black lines indicate medians. Statistical analysis: unpaired, two-tailed Student's *t*-test on biological replicate data. $n = 3$ per genotype, 703 plaques for 5×FAD, 846 plaques for *Cnp*$^{-/-}$;5×FAD in total.

Source data

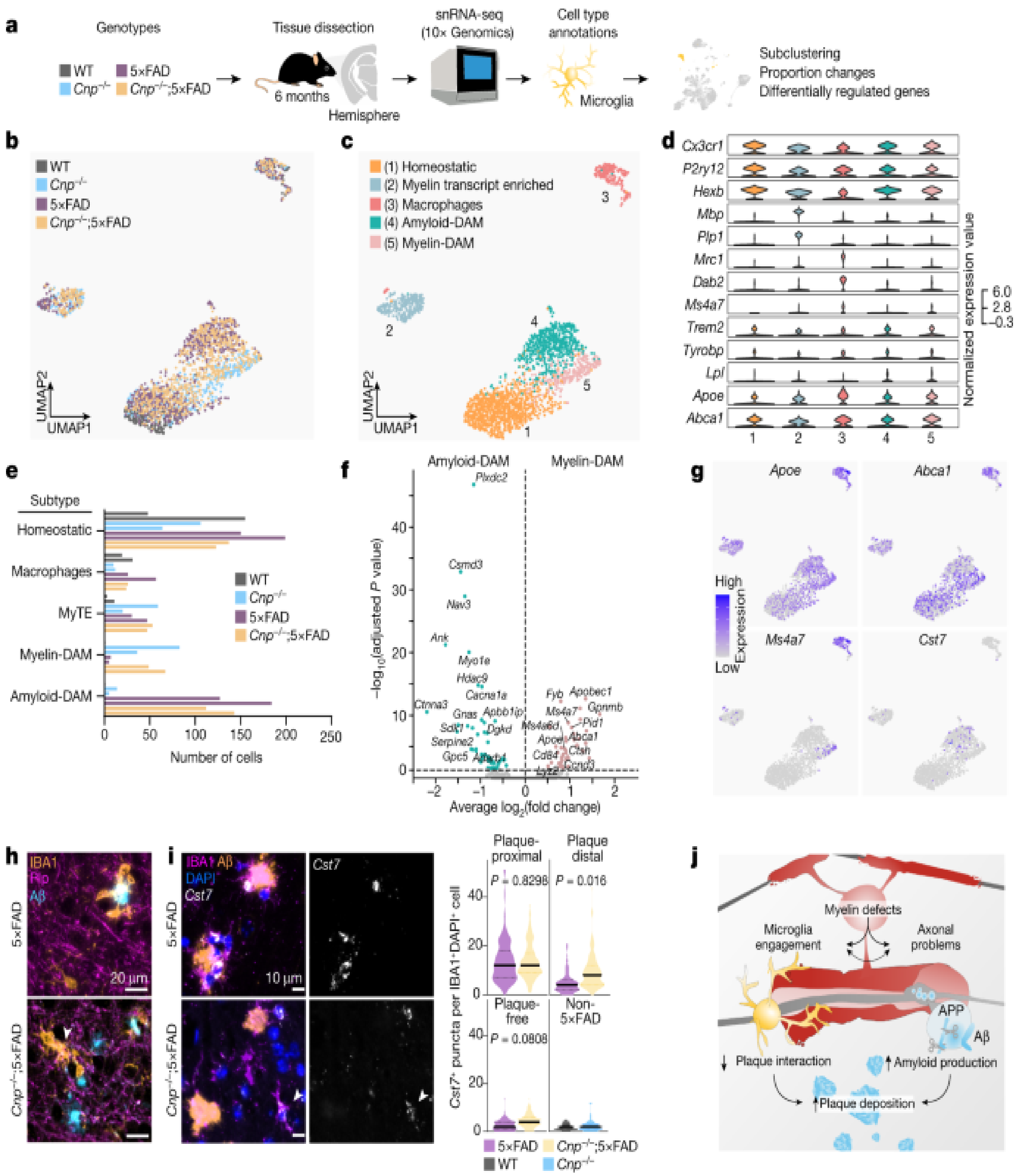
Similar defects in microglial plaque-corralling have been previously described in *Trem2* loss-of-function scenarios[34,35,36]. We therefore investigated whether myelin dysfunction interferes with the upregulation of *Trem2* and the induction of the disease-associated microglia (DAM) signature that is associated with a plaque-corralling phenotype[37]. To this end, we performed magnetic activated cell sorting (MACS) of microglia and bulk RNA sequencing (RNA-seq) (Fig. 4c). Principal component analysis (PCA) (Fig. 4d and Extended Data Fig. 8b) and differential gene expression analysis revealed incremental downregulation of homeostasis markers throughout the WT, 5×FAD, *Cnp*$^{-/-}$ and *Cnp*$^{-/-}$;5×FAD trajectory with concomitant increases in activation and inflammatory genes (Fig. 4e,f and

Extended Data Fig. 8a–e). Clustering of differentially expressed genes (DEGs) based on normalized expression and subsequent gene ontology (GO) enrichment analysis confirmed an overall more inflammatory profile of microglia isolated from *Cnp*$^{-/-}$ mice (Extended Data Fig. 8d,e and Supplementary Tables 1 and 2). Classical DAM signature genes were upregulated in microglia from 5×FAD mice and were further increased in microglia from *Cnp*$^{-/-}$ mice and *Cnp*$^{-/-}$;5×FAD mice (Fig. 4e,f and Extended Data Fig. 8c). DEG analysis between *Cnp*$^{-/-}$;5×FAD mice and 5×FAD mice identified many previously reported DAM signature genes (*Clec7a*, *Gpnmb*, *Apoe*, *Spp1*, *Axl* and *Itgax*) and genes of the *Ms4a* cluster as highly upregulated in *Cnp*$^{-/-}$;5×FAD mice (Extended Data Fig. 8f–h). Such a signature was previously described to be a feature of highly phagocytic microglia in early development and for brain border macrophages[38]. *Trem2* and *Tyrobp*, however, showed unaltered induction levels (Fig. 4f). As *Ms4a* genes can regulate the level of soluble TREM2 and modify AD risk[39], we used immunoblot analysis to rule out the possibility that myelin dysfunction affects TREM2 protein levels or cleavage through increased *Ms4a* expression (Extended Data Fig. 9a). Functional enrichment analysis revealed upregulation of GO terms and pathways related to the immune response in *Cnp*$^{-/-}$;5×FAD mice (Extended Data Fig. 8i). STRING (Extended Data Fig. 8i) and DEG analyses showed substantial upregulation of lipid-metabolism-related genes (*Apoe*, *Apoc1*, *Apoc4* and *Lpl*) in *Cnp*$^{-/-}$;5×FAD mice (Extended Data Fig. 8j). Results of co-immunostaining experiments showed increased levels of APOE, a well-known factor in Aβ aggregation[40], in microglia (Fig. 4g) and amyloid plaques in *Cnp*$^{-/-}$;5×FAD mice (Fig. 4h).

To better understand the DAM-like microglial signature induced by myelin defects, we further performed single-nuclei RNA-seq (snRNA-seq) on brain tissue from *Cnp*$^{-/-}$;5×FAD mice and respective controls (Fig. 5a and Extended Data Fig. 9b–d). Cluster analysis identified five major microglia and macrophage subpopulations, including two distinct clusters (clusters 4 and 5) with high expression of DAM marker genes[37] (*Trem2*, *Lpl* and *Spp1*) (Fig. 5b–d). Cluster 4 was almost exclusively derived from cells with the 5×FAD genotype, whereas cluster 5 consisted of cells from the *Cnp*$^{-/-}$ background (Fig. 5e). We therefore termed these two clusters

myelin-DAM and amyloid-DAM, respectively. Differential expression analysis between myelin-DAM and amyloid-DAM revealed upregulation of lipid-metabolism-related genes (*Apoe*, *Abca1* and *Apobec1*) and genes of the *Ms4a* cluster in myelin-DAM, which corroborated findings from our bulk RNA-seq experiment (Fig. 5f and Extended Data Fig. 9e). Genes specifically upregulated in amyloid-DAM included *Cst3*, *Ctnna3* (which encodes alpha-T-catenin) and *Gpc5* (which encodes the heparan sulfate proteoglycan glypican 5). Despite a low detection rate, *Cst7* was also specifically enriched in amyloid-DAM (Fig. 5g). Notably, when comparing microglia from *Cnp*$^{-/-}$;5×FAD mice with 5×FAD mice within the amyloid-DAM cluster, we still observed upregulation of *Apoe in Cnp*$^{-/-}$;5×FAD microglia (Extended Data Fig. 9g,h). Analyses of DEGs shared between DAM clusters and genotype comparisons within DAM clusters revealed that both amyloid-DAM and myelin-DAM of the *Cnp*$^{-/-}$;5×FAD crossbred mice retained signature features of the other DAM cluster, which highlights the intermediate phenotype of microglia in this mouse model (Extended Data Fig. 10h,i).

Fig. 5: Myelin dysfunction induces a DAM-like state distinct to amyloid deposition as determined by snRNA-seq.



a, Experimental setup for studying microglia states associated with myelin disease (myelin-DAM) and amyloid-disease (amyloid-DAM) and in combination. Brain hemispheres were isolated from 6-month-old WT, *Cnp*$^{-/-}$, 5×FAD and *Cnp*$^{-/-}$;5×FAD mice and subjected to snRNA-seq. Cell types were identified on the basis of marker genes, and microglia were subset for further analysis. **b**,**c**, Uniform manifold approximation and projection (UMAP) visualization of microglia subsets coloured by genotypes (**b**) or

subpopulations (**c**). **d**, Violin plots showing expression of microglia subpopulation marker genes. **e**, Bar plot showing the distribution of microglia subpopulations across genotypes. MyTE, myelin transcripts enriched. **f**, Differentially regulated genes between amyloid-DAM and myelin-DAM. **g**, Feature plots showing expression of *Apoe*, *Abca1*, *Ms4a7* and *Cst7* in microglia subpopulations. **h**, Representative example images of microglia distraction in *Cnp*$^{-/-}$;5×FAD mice. The arrowhead in the bottom panel highlights an activated microglial cell that is engaged in myelin phagocytosis and does not react to nearby plaques. The same observations were made in five independent samples per mice per group. **i**, Left, amyloid and microglia immunostaining and single-molecule fluorescence in situ hybridization for the amyloid-DAM marker *Cst7*. Right, violin plots show the amount of *Cst7* signal per microglia. Microglia were separated into groups according to their location in relation to amyloid plaques (plaque-proximal, plaque-distant and plaque-free regions). The following number of microglia were analysed: 5×FAD: plaque-proximal = 161, plaque-distant = 111, plaque-free = 128; *Cnp*$^{-/-}$;5×FAD: plaque-proximal = 136, plaque-distant = 184, plaque-free = 184 from 3 mice per replicates per group ($n = 3$). For *Cnp*$^{-/-}$, $n = 61$ and for WT, $n = 63$ from one animal each. Black lines indicate medians. Statistical analysis: two-sided, unpaired Student's t-test (P values are indicated in the graphs) on biological replicate data. **j**, Scheme illustrating how myelin dysfunction acts as an AD risk factor. Upstream myelin defects cause microglia engagement and axonal transport problems. The two downstream pathologies are probably interrelated (dashed arrow). Axonal problems lead to endosome and lysosome accumulation and enhanced amyloid production. Simultaneously, microglia become increasingly engaged with defective myelin, which reduces their interaction with amyloid plaques. Both processes contribute to the enhanced deposition of amyloid.

Source data

Overall, our single-nucleus transcriptome analysis indicated that microglia respond in a similar but distinct manner to myelin dysfunction and the amyloid plaques they encounter. In the *Cnp*$^{-/-}$;5×FAD model, in which the two pathologies co-exist, both microglia states are adequately induced (despite modifications to the signatures). Our imaging data, by contrast,

showed an almost complete loss of plaque-corralling microglia, which were previously identified as the in situ equivalent to DAM[37]. We wondered whether myelin damage would engage a large fraction of microglia in myelin clearance, thereby distracting even successfully induced amyloid-DAM away from plaques. Immunostaining for three markers showed that a large number of microglia and plaque-diverted microglia in *Cnp*$^{-/-}$;5×FAD mice seemed to be reactively engaged in myelin phagocytosis (Fig. 5h and Extended Data Fig. 9j). *Cst7* in situ hybridization—a marker for amyloid-DAM—showed that amyloid-DAM in *Cnp*$^{-/-}$;5×FAD mice were frequently found distal to the plaque, which was in contrast to 5×FAD control mice (Fig. 5i). We propose that such distraction of microglia leads to a faster build-up of amyloid in the brain. Moreover, factors secreted by activated but myelin-engaged microglia (such as APOE) may further fuel plaque seeding.

We next investigated whether altered microglia phenotypes also contribute to amelioration of plaque burden in forebrain shiverer;5×FAD mice (which lack myelin in the forebrain). To this end, we performed snRNA-seq analysis of microglia states in forebrain shiverer mice and contrasted those with the microglia signature found in *Cnp*$^{-/-}$ animals at 3 months of age (Extended Data Fig. 10a). In forebrain shiverer mice, we failed to detect activated microglia or myelin-DAM compared with microglia from *Cnp*$^{-/-}$ mice (Extended Data Fig. 10b–d). Similarly, immunostaining showed that microglia from forebrain shiverer;5×FAD mice remained in an homeostatic state and displayed a highly ramified morphology. However, we detected an increase in microglia cell numbers (Extended Data Fig. 10e). This could convey a certain protection against plaque build-up early in disease as more microglia could potentially clear amyloid. Also, contrary to mice with dysfunctional myelin, the lack of myelin in forebrain shiverer;5×FAD mice did not lead to axonal swellings or alterations in APP metabolism (Extended Data Fig. 10f,g).

Our findings and experimental models (myelin defects on top of an AD background) are highly relevant to the human disease as they mimic the state of the old ‘comorbid’ brain. That is, microglia would see multiple pathological stimuli simultaneously—such as age-related myelin breakdown and amyloid—similar to microglia in *Cnp*$^{-/-}$5×FAD mice. A re-analysis of published human AD snRNA-seq datasets showed that activated microglia

—in contrast to pure amyloid-DAM in 5×FAD mice—showed upregulation of *MS4A* cluster genes, thereby demonstrating a myelin clearance-related signature (Extended Data Fig. 10h–o).

Ageing myelin as a risk factor for AD

Only a few studies have experimentally investigated oligodendrocytes and myelin in amyloid mouse models of AD[20,21,41,42,43]. Nearly all have focused on the effect of amyloid plaques and Aβ on oligodendrocytes and myelin, partially describing secondary demyelination or hypomyelination. Similarly, single cell RNA-seq analysis of mouse models of AD demonstrated signature alterations in oligodendrocytes, specifically nearby amyloid plaques[44,45]. RNA-seq studies using human AD autopsy samples have identified myelin-related transcripts among the top altered gene clusters in patients with AD[35,46]. These findings, however, probably reflect the downstream effects of overt Aβ and tau pathology on oligodendrocytes in the latest stages of disease. In our study, we specifically asked whether myelin defects are upstream of amyloid deposition.

Based on our findings, we propose a working model for AD (Fig. 5j) in which myelin dysfunction in the ageing forebrain (modelled by the premature ageing of white matter in myelin mutant mice) causes microglia to become engaged. This microglial activity interferes with the ability to clear Aβ deposits and prevent plaque formation. Simultaneously, ageing myelin loses its axon-supportive functions, leading to axonal distress, which in turn increases neuronal BACE1 and APP CTF levels suggestive of enhanced Aβ production. It is tempting to speculate that other well-known AD risk factors, such as traumatic brain injury or cardiovascular diseases, also convey risk by negatively affecting myelin health or at least through similar mechanisms (that is, distracting microglia).

The identified plaque-promoting factors, axonal distress and microgliosis are tightly linked downstream phenomena of myelin injury, which makes it difficult to study them in isolation. We have previously shown that depletion of microglia in *Cnp*$^{-/-}$ mice can reduce the number of axonal swellings[47], which indicates that axonal distress and APP metabolism changes could be downstream of microglia activation. However, nonspecific microglia

activation induced by lipopolysaccharide does not have the same amyloid-promoting effect as myelin dysfunction, as increases, decreases or no changes in amyloid burden have been reported[48,49,50,51]. By contrast, myelin defects seem to be unique drivers of microglia activation in the ageing brain (as previously shown[3]), which challenge microglia with the uptake and digestion of large amounts of lipids and highly compacted membranes.

An important clinical implication of our findings is the co-existence of AD and multiple sclerosis (MS). Based on our findings in EAE and cuprizone-treated AD mice, we would suspect an increased rate of AD as a comorbidity of MS. A recent study[52] has shown that patients with MS have a higher risk of receiving an AD or dementia diagnosis. However, given the very limited dataset, further epidemiological studies concerning the comorbidity of MS and AD are needed[53].

A role of myelin deterioration in the progression of AD has been previously proposed in light of correlative macroscopic-scale brain imaging studies[54], but they lacked any evidence of causality. Our experimental findings in genetic models of AD now provide a causal link with a molecular underpinning and are in line with the prevailing amyloid hypothesis and the role of neuroinflammation in disease progression. Moreover, we can position the well-documented loss of myelin integrity in the ageing primate brain as an upstream initiator of AD pathology, which might help to explain why age is the major AD risk factor. Our working model is further supported by previous classical neuropathological observations[55], in which a conspicuous temporal relationship with CNS myelination was reported, which occurs latest in those brain regions that are the first to develop AD pathology. Mechanistically, cortical ensheathment by oligodendrocytes is thin and may lack the distinct myelinic channel architecture important for glial metabolic support found in subcortical tracts that appear resistant to AD pathology.

Our discovery that age-dependent loss of myelin integrity can be a driver or risk factor of amyloid deposition changes our view on the role of oligodendrocytes in AD—from passive bystanders to active contributors of disease progression. If further corroborated in humans, promoting myelin health should be considered as a therapeutic target for the delay or prevention of AD.

Methods

Human tissue analysis

Selection of patients was performed from a pool of approximately 400 individuals, in which an autopsy with neuropathological evaluation was performed between 2018 and 2019 as a matter of routine procedure following death at the Leipzig University Hospital. Ethics oversight was performed by the Ethics Board of the Leipzig University Hospital. In the individual contracts that govern medical treatment, all patients included in the study provided written consent following hospital admission to the scientific use of tissue removed and stored after any biopsy or during autopsy.

Selection of patients was performed according to exclusion and inclusion criteria. Samples were anonymized and processed in a blinded manner. Selected patients were of mixed age and sex, aged between >60 years and <90 years of age, had a clinical history of dementia and a National Institute on Aging–Alzheimer's Association (NIA–AA) score in neuropathological assessment between A2B3C2 and A3B2C2 (inclusion criteria) and did not suffer from another severe neurological disorder (exclusion criteria). In addition, individuals of the same age range without any clinical or neuropathological record of neurological disease were selected. No other criteria besides the described characteristics were applied. In total, paraffin-embedded CNS tissue from three patients with moderate to pronounced AD neuropathological changes according to the NIA–AA (see above) and from three unaffected individuals were histologically evaluated. In our analysis, we used biopsy samples from the medial temporal lobe containing the hippocampal formation. We performed histological assessment for intracortical myelin content on the human tissue provided. For this, we sectioned paraffin blocks (5 μm) and stained paraffin sections for CNP, PLP, IBA1 and amyloid plaques simultaneously (see the section 'Paraffin slices and histological stainings' for details). Quantification of microgliosis (IBA1) and myelin levels (CNP and PLP) were performed in the trans-entorhinal cortex by performing thresholding, and the percentage of positive area was determined using Fiji in equally sized ROIs in the transentorhinal cortex.

Mouse strains and husbandry

Animal experiments were conducted in compliance with German animal welfare practices and approved by the local authorities (Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Niedersachsen). Mice were group-housed in the local animal facility of the Max Planck Institute for Multidisciplinary Sciences under a 12-h dark–12-h light cycle and fed ad libitum. Both sexes were used throughout the study. The following original mouse strains were used to generate crossbreedings: 5×FAD (ref. [56]), APP^{NLGF} (ref. [57]), $Cnp^{-/-}$ (ref. [58]), $Plp^{-/y}$ (ref. [59]), $Plp^{fl/fl}$ (ref. [60]), $Mbp^{fl/fl}$ (ref. [61]), Emx^{cre} (ref. [62]) and $Foxg1^{cre}$ (ref. [63]). For analysis, either littermate controls were used or a corresponding control line from the initial F_1 generation was generated. The age of animals is given in the respective figure legend. All animals were maintained on a C57BL/6 background. Genotyping was performed on clips derived from earmarking according to standard protocols (see references for original mouse strains). The genotype was confirmed by regenotyping on a tail biopsy following euthanasia of the animal at the end of the respective experiment.

Demyelination models

As demyelination models, we used cuprizone and EAE, and experiments were conducted as previously described[64]. For cuprizone-mediated demyelination, male 14-week-old 5×FAD mice were fed a powder diet containing 0.2% (w/w) cuprizone (Sigma-Aldrich) for 4 weeks followed by a 4-week recovery period on normal pelleted food without cuprizone supplementation. Control 5×FAD mice received a standard diet without cuprizone. Animals were perfused at 18 weeks of age, and brain tissue was analysed by LSM and epifluorescence microscopy. For EAE experiments, 10-week-old 5×FAD mice were immunized against myelin oligodendrocyte glycoprotein (MOG) by subcutaneously injecting 200 mg MOG(33–35) peptide in complete Freund's adjuvant (*Mycobacterium tuberculosis* at 3.75 mg ml^{-1}; BD) followed by injecting 500 ng of pertussis toxin (Sigma) at day 1 and 3 after EAE induction (d.p.i. 1 and 3). Animals were checked on a daily basis, and a neurological disease score was determined according to the following parameters: 0, normal; 0.5, loss of tail tip tone; 1, loss of tail tone; 1.5, ataxia and mild walking deficits (slip off the grid); 2, mild hind

limb weakness, severe gait ataxia and twist of the tail causing rotation of the whole body; 2.5, moderate hind limb weakness and inability to grip the grid with the hind paw but ability to stay on an upright tilted grid; 3, mild paraparesis and falls from an upright tilted grid; 3.5, paraparesis of the hind limbs (legs strongly affected but move clearly); 4, paralysis of the hind limbs and weakness in the forelimbs; 4.5, forelimbs paralysed; 5, moribund or dead. Animals were perfused at 28 d.p.i. at 14 weeks of age. All immunized 5×FAD animals developed EAE. Brain and spinal cords were analysed by LSM and epifluorescence microscopy.

Mouse behavioural testing

Hindlimb clasping was assessed by suspending the mouse on its tail for 5 s, carefully observing movement of the hindlimbs and scoring movement impairments according to a score from 0–4 as follows: 0, no impairment, hindlimbs normally spread and moving; 1, one hindlimb shows slightly less mobility; 2, both hindlimbs show less mobility; 3, both hindlimbs show reduced mobility and reduced spread; 4, both hindlimbs show severely reduced movement and severely reduced spread. Videos were recorded. Animals were tested in the elevated plus (EPM) and Y-maze (YM) on consecutive days. In the EPM experiment, animals were allowed to freely explore an elevated cross-shaped platform with two opposing enclosed arms and two opposing open arms for 5 min without the experimenter present. In the YM experiment, animals were allowed to explore a Y-shaped maze with enclosed arms for 8 min without human interference. Videos were recorded and analysed using the Bioobserve Viewer behavioural analysis setup with automated tracking. Zones (open arms, closed arms, centre) and object detection settings were optimized according to the maze type used. After the run, positional data, track length and full track images were exported. For the EPM experiment, the time spent in both open arms was summed and plotted. For the YM experiment, the order of arm entries was recorded, and the number of correct triads (consecutive visit of the three different arms) determined. The alternation index was calculated according to the formula alternation index = number trials/number of arm entries – 2. For a successful visit to an arm of the maze, animals had to enter an arm with their full body (excluding the tail).

For statistical analysis of the behavioural data, we report the results of several different type 3 analysis of variance (ANOVA) tests that calculated the main effects for the 5×FAD mice and myelin mutant mice as well as their interaction. All analyses were conducted in R (v.4.04). The ANOVA tests were computed using the afex package (afex v.0.28-1)[65]. Post hoc testing was performed using Tukey's multiple comparison test in Prism 8.

Tissue preparation

For microscopy analysis, animals were deeply anaesthetized or euthanised by CO_2 asphyxiation and subsequently transcardially perfused with ice-cold Hank's buffered salt solution and 4% paraformaldehyde (PFA) in 0.1 M phosphate buffer, pH 7.4. Brains and spinal cords were removed and post-fixed overnight in 4% PFA and phosphate buffer. Brains were washed once in PBS pH 7.4 and stored in PBS at 4 °C until further use. For biochemical analysis of full brains, animals were killed by cervical dislocation, brain and spinal cord were extracted and fresh-frozen on dry ice. Tissue was stored at −80 °C until further use. For microdissection of subcortical white matter and cortical tissue, animals were killed by cervical dislocation and the brain was quickly removed and submerged in ice-cold Hank's buffered salt solution. The brain was inserted into a custom-made brain matrix and the brain was sliced coronally in about 1-mm-thick slices. Brain slices were spread onto an ice-cold glass plate, and cortex and subcortical white matter were separated and excised from each individual brain slices. Tissue was immediately frozen on dry ice and stored at −80 °C until further use.

Whole tissue staining and clearing for LSM

Fixed hemibrains and spinal cords were pretreated and permeabilized following a modified iDISCO protocol[66]. In brief, tissue samples were dehydrated with an ascending concentration of methanol in PBS (50% once, 80% once, 100% twice, 1 h each). Tissue samples were then bleached with a 1:1:4 ratio of H_2O_2:DMSO:methanol overnight at 4 °C. Further dehydration was followed by 100% methanol incubation at 4 °C (30 min), −20 °C (3 h) and overnight storage at 4 °C. Samples were incubated the following day in methanol with 20% DMSO before subjecting them to gradual rehydration in a descending methanol in PBS series (80% once, 50% once, 0% once, 1 h

each). We then washed the tissue samples with a detergent mixture of 0.2% Triton X-100 in PBS (twice, 1 h) and permeabilized them overnight at 37 °C in PBS with 0.2% Triton X-100, 20% DMSO and 0.3 M glycine. After permeabilization, tissue samples were either stained with the Congo red dye (Sigma Aldrich) for β-sheet structures within amyloid plaques or immunolabelled with antibodies of interest. For Congo red staining, tissue samples were washed with PBS and 0.2% Tween-20, 10 mg ml^{-1} heparin, 5 mM sodium azide (PTwH; twice, 1 h) before immersing them for 3 days at 37 °C in 0.005% w/v Congo red (100× stock solution in 50% ethanol). For immunolabelling, following tissue permeabilization and glycine treatment, samples were blocked in PBS with 0.2% Triton X-100, 10% DMSO and 6% goat serum (GS) for 3 days followed by washing in PTwH (twice, 1 h) and incubation in primary antibodies with the appropriate dilution factors (1:500, rabbit, anti-IBA1, Wako) in PTwH with 0.2% Triton X-100, 5% DMSO and 3% GS for 14 days at 37 °C. Following completion of labelling with primary antibody, tissue samples were washed in PTwH (6 times, 10 min) and stored overnight at 37 °C. For secondary antibody labelling, tissue samples were again incubated in secondary antibodies with the appropriate dilution factors (anti-rabbit Dylight 633, 1:500, Thermo-Fisher) in PTwH with 3% GS for 7 days at 37 °C. Before clearing, spinal cords were fixed in 1.5% w/v Phytagel in water. Dyed or immunolabelled tissue samples were washed in PTwH (3 times, 10 min each) before rehydration through an ascending concentration of methanol in PBS (20% once, 40% once, 60% once, 80% once, 100% once, 1 h each) and delipidation in a 1:2 mixture of methanol:dichloromethane (once, 1 h 40 min). Last, samples were cleared by immersing them in ethyl cinnamate (Eci, Sigma-Aldrich) until transparent. All incubation steps were carried out at constant medium-speed rotation at the indicated temperatures. Samples were stored at room temperature in Eci until imaging.

In toto imaging of whole tissues and analysis and visualization

Cleared hemibrains and spinal cords were imaged in toto with a light sheet microscope setup (UltraMicroscope II, LaVision Biotec) equipped with a ×2 objective lens, a zoom body and a corrected dipping cap. Samples were imaged submerged in a sample chamber containing Eci. For all hemibrain imaging, hemibrains were mounted medial-side down on the sample holder

to acquire sagittal images. Images were acquired using ImspectorPro (v.7.124, LaVision Biotech) software in the mosaic acquisition mode with the following settings: 5 µm light sheet thickness, 20% sheet width, 0.154 sheet numerical aperture, 4 µm *z*-step size, 1,000 × 1,600 pixels field of view, 4 × 4 tiling, dual light sheet illumination and 100 ms camera exposure time. Red fluorescence was recorded with 561 nm laser excitation at 80% laser power and a 585/40 nm emission filter. Far-red fluorescence was recorded with 640 nm laser excitation at 30% laser power and a 680/30 nm emission filter. Autofluorescence was recorded with 488 nm laser excitation at 50% and 525/20 nm emission filter. Images were imported into Vision4D (v.3.2; Arivis) and stitched using the tile sorter setup. For some, images were imported and stitched using Imaris Importer and Stitcher (v.9.1; Bitplane). Rendered whole hemibrains were then processed, and three main ROIs were manually annotated based on the sagittal Allen mouse brain atlas, namely the isocortex, the hippocampus and the alveus. All ROIs were first manually traced in 2D planes to automatically extrapolate the 3D ROIs. Cortical and hippocampal annotations were cropped with a medial cutoff of approximately 0.4 mm and a lateral cutoff of 4.4 mm, which would span the dorsal isocortex and the entire hippocampal formation of one hemibrain. The lateral cutoff for the alveus ROI is the plane where the hippocampal formation appears in 2D. Next, we segmented amyloid plaques within the ROIs. For 3-month-old 5×FAD brain data, we typically used automated intensity thresholding. For 6-month-old 5×FAD data, plaques were segmented using the blob finder algorithm in Vision4D with the following parameters: 20 µm object size, 10–15% probability threshold and 0% split sensitivity. Once segmentation was performed, stringent noise removal was performed by deleting objects with voxel sizes <10 from the object table. Objects were then carefully reviewed, and any additional noise, which might include but are not restricted to blood vessels and nuclei, were manually removed from the dataset. For quantification of amyloid plaques in the APP^{NLGF} brains that typically stained much weaker for Congo red, plaque segmentation was performed using the machine-learning segmenter in Vision4D. Object parameters and ROI volumes were extracted for further quantification. For plaque visualization in the different ROIs, objects are represented in centroids and colour-coded according to location. The following different plaque burden parameters were assessed: plaque count (number of plaques in ROI), total plaque volume (sum of all individual

plaque volumes), plaque size (average volume of plaques), plaque density (number of plaques normalized to ROI volume), percentage volume (ROI volume covered by sum of plaque volume).

Paraffin slices and histological stainings

Fixated hemibrains and spinal cords were subjected to dehydration steps (50% ethanol, 80% ethanol, 100% ethanol, 100% isopropanol, 50% isopropanol and 50% xylol, twice 100% xylol) followed by paraffinization on a STP 120 tissue processing machine (Leica Microsystems). Samples were retrieved and embedded in paraffin blocks on a HistoStar embedding workstation (Epredia). Paraffin-embedded blocks were sectioned coronally, whereas spinal cords were sectioned longitudinally at 5-µm slice thickness, slices were mounted onto slides and dried overnight. Slides were deparaffinized at 60 °C followed by incubation in xylol (100% twice) and a 1:1 mixture of xylol and isopropanol (once) for 10 min each. The slides were rehydrated in a descending ethanol series. This was followed by incubation in either acidic antigen retrieval solution (pH 6.0) or basic antigen retrieval solution (10 mM Tris and 1 mM EDTA pH 10) for 5 min and boiling for 10 min. The samples were cooled for 20 min and washed in distilled water for 1 min before subsequent permeabilization in 0.1% Triton X-100 in PBS. For Aβ antibody staining, samples were subjected to an extra antigen retrieval step in 88% formic acid to loosen the β-sheet structure for optimal antibody binding for 3 min. For plaque ApoE staining, formic acid treatment was prolonged (10 min, fresh formic acid). This was followed by washes in PBS (twice, 5 min) and blocking with 10% GS in PBS for 1 h at room temperature. For chromogenic labelling, an additional step of inactivation of endogenous peroxidases was implemented before blocking by incubation in 3% H_2O_2. After blocking, slices were incubated in primary antibody solution (PBS, 10% GS or PBS, 5% BSA) overnight at 4 °C in coverplates (Epredia). The following antibodies were used in this study: anti-IBA1 (rabbit, Wako; 1:1,000); anti-Aβ-6E10 (mouse, BioLegend; 1:1,000); anti-CNP (mouse, AMAb91072, Atlas; 1:1,000); anti-PLP-clone aa3 (rat, culture supernatant; 1:200); anti-BACE1 (rabbit, ab183612, Abcam; 1:100); anti-MBP (rabbit, serum, custom-made in the Nave Laboratory; 1:1,000); anti-GFAP (mouse, GA5, Leica; 1:200); anti-N-terminal APP (22c11, Merck; 1:1,000); anti-C-terminal APP (rabbit, 127-003, Synaptic Systems; 1:1,000);

anti-A$\beta_{x–42}$-D3E10 (rabbit, Cell Signaling Technology; 1:1,000); anti-PSEN2 (rabbit, Abcam; 1:100); anti-ApoE D7I9N (rabbit, Cell Signaling Technology; 1:500); anti-ApoE clone 26c11 (mouse, 1:500, provided by C. Haass Laboratory); sAPPβ_{swe} clone 6A1 (mouse, IBL; 1:1,000); anti-A$\beta_{1–x}$ (mouse, 80C2, Synaptic Signalling; 1:200); anti-A$\beta_{1–x}$ (mouse, Moab2-6C3, Abcam; 1:200); anti-A$\beta_{4–x}$ (guinea pig, 029-2 (ref. [67]), custom-made by O.W.; 1:200); and anti-APP/A$\beta_{17–24}$ clone 4G8 (mouse, BioLegend; 1:1,000). For immunofluorescence staining, samples were washed in PBS or Tris and 2% milk and incubated with the corresponding fluorescent secondary antibody diluted in PBS containing 10% GS for 2 h at room temperature in the dark. The following fluorescently conjugated secondary antibodies were used: anti-mouse Alexa555 (donkey/goat, Thermo-Fisher; 1:1,000); anti-mouse Dylight633 (donkey/goat, Thermo-Fisher; 1:1,000); anti-rabbit Alexa555 (donkey/goat, Thermo-Fisher; 1:1,000); and anti-rabbit Dylight633 (donkey/goat, Thermo-Fisher; 1:1,000). Amyloid plaques were stained using the β-sheet dye methoxy-X04 (4 μg ml^{-1} in 50% ethanol, Tocris) for 30 min at room temperature and contrasting in 50% ethanol. Nuclei were stained with either DAPI (300 nM, Thermo-Fisher) or ToPro3 (1:1,000, Thermo-Fisher) in PBS for 5 min at room temperature. Slides were again washed in PBS and mounted with Aqua PolyMount mounting medium (PolySciences). For chromogenic labelling, a LSAB2 kit (Dako) for rabbit/mouse and a DAB-Kit from Zytomed was used according to the manufacturers' protocols. Slides were then rinsed in water, dehydrated and mounted using Eukitt (Sigma-Aldrich).

In situ hybridization

We performed RNAscope Fluorescent Multiplex assays (ACD Bio) according to the manufacturer's instructions for both paraffin-embedded tissue and cryo-tissue sections. In brief, paraffin-fixed slices underwent pretreatment consisting of deparaffinization steps at 60 °C for 1 h, followed by incubation in 100% xylol (5 min, twice) and 100% ethanol (2 min, twice). Slides were then incubated with hydrogen peroxide at 40 °C for 15 min. After a wash with dH_2O, slides were boiled in the target retrieval reagent for 10 min and washed with dH_2O (15 s, once) and with 100% ethanol (3 min, once). After drying, a hydrophobic barrier was drawn and protease digestion

was performed by incubating slides with RNAscope Protease Plus for 15 min at 40 °C. Before applying the probes, slides were washed with dH_2O (2 min, twice). The following RNAscope probes were used: Mm-CST7-C3 (498711-C3) and Mm-Ms4a7-C2 (314601-C2). The probe was hybridized for 2 h at 40 °C. Slides were stored overnight in a 5×SSC buffer. The following day, signal amplification was performed by applying amplification reagents provided in the kit. This was followed by incubation with HRP corresponding to the channel probe for 15 min at 40 °C and visualization was performed using the OpalTM 690 fluorophore for 30 min at 40 °C. After the final wash, we proceeded with immunohistochemical co-staining. Blocking of the slides was done with 10% GS in PBS for 1 h at room temperature. After blocking, incubation in primary antibody solution (PBS and 10% GS) was done overnight at 4 °C. The following primary antibodies were used: anti-IBA1 (rabbit, Wako; 1:1,000) and anti-Aβ-6E10 (mouse, BioLegend; 1:1,000). *Ms4a7* RNAscope was performed on cryo-sections with IBA1 co-labelling. *Cst7* RNAscope was performed on paraffin sections with 6E10 and IBA1 co-labelling. Slides were washed in PBS (5 min, 3 times) and labelled with corresponding fluorescent secondary antibody solution (PBS and 10% GS) for 2 h at room temperature. The following fluorescently conjugated secondary antibodies were used: anti-mouse Dylight488 (donkey, Thermo-Fisher; 1:1,000) and anti-rabbit Dylight555 (donkey, Thermo-Fisher; 1:1,000). Finally, slides were washed in PBS (5 min, 3 times) and mounted with Aqua PolyMount mounting medium (PolySciences). Imaging was performed using a Zeiss Observer Z1 microscope equipped with Plan-Apochromat ×20/08 objective. Quantification of *Cst7* RNAscope was performed by identifying plaque-corralling microglia and plaque-distant microglia (without obvious plaque interaction), and the amount of *Cst7* puncta within these microglia was counted.

Epifluorescence and brightfield microscopy

Epifluorescence microscopy was performed on a Zeiss Observer Z1 microscope equipped with Plan-Apochromat ×20/08 and Fluar ×2.5/0.12 objectives, a Colibri 5 LED light source (630 nm, 555 nm, 475 nm and 385 nm excitation), 96 HE BFP, 90 HE DAPI, GFP, Cy3, Cy5, 38 GFP, 43 DsRed, 50 Cy5 Zeiss filter sets, a Axiocam MrM and a SMC900 motorized

stage. For whole-brain-slice microscopy, a preview scan at ×2.5 magnification was taken, and focus support points were distributed and manually set for imaging at ×20 magnification in ZEN imaging software (ZEN 2011 v.2.0, Zeiss). Tiled images were stitched in ZEN. For visualization, pseudocolours (cyan, magenta, yellow) were assigned to different channels, the intensity was adjusted and images were exported into ZEN. Brightfield microscopy of DAB-stained slices was performed on a Zeiss Axiophot Imager.Z1 equipped with a Achroplan ×4/0.1, PlanFluar ×20/0.75 and Plan Neofluar ×40/0.75 objectives and a AxioCamMrc camera. For whole-brain-slice microscopy, a preview scan at ×4 magnification was taken, and focus support points were distributed and manually set for imaging at ×40 magnification in ZEN imaging software (Zeiss). Tiled images were stitched in ZEN. Brightness and contrast of RGB images were adjusted and images exported into ZEN.

2D microscopy quantification

2D image analysis was performed in Fiji (v.1.53c)[68] and in ZEN. For quantification of amyloid load, thresholding was performed to segment Aβ deposits (either stained by Me-04 or Aβ plaques), and the positive area was calculated in the respective ROI. For analysis of the plaque-corralling phenotype in the 5×FAD model, in IBA1^{+}Aβ^{+} co-stainings, individual plaques were segmented using thresholding, and the IBA1^{+} area in each individual plaque was calculated using a Fiji macro-script. IBA1 coverage was expressed as a percentage. As plaques in the *APP*^{NLGF} line are not as compacted as plaques in the 5×FAD model and therefore commonly appear ‘fragmented’, the same type of analysis could not be reliably performed in *APP*^{NLGF} mice (overestimation of single plaque structures). For both 5×FAD mice and *APP*^{NLGF} mice, the number of plaque-associated microglia per plaque (50 representative plaques with diameter >20 µm for each animal) was manually counted. For quantification of ApoE levels in plaques, ApoE^{+} plaques were segmented, and the raw mean fluorescence per plaque was calculated. For quantification of microgliosis in *Cnp*$^{-/-}$;5×FAD, *Plp*$^{-/y}$;5×FAD and *Emx*^{cre}*Mbp*$^{fl/fl}$ experimental cohorts, the amount of IBA1^{+}DAPI^{+} cells in the ROI (corpus callosum or cortex) was counted and normalized to the area examined. Additionally, the IBA1^{+} area fraction was

determined in Fiji by performing thresholding. The number of axonal swellings positive for APP-processing, APP and Aβ antibodies was determined by counting these structures in the respective ROI (fimbria) and normalization to the area examined.

EM analyses

Fixed brain tissues were immersed in ice-cold PBS and cut into a coronal section with 300 µm thickness using a vibratome (Leica VT1200). Tissue containing the medial corpus callosum and cingulate cortex with adjacent tissue was punched with a 2 mm diameter punching tool and embedded in Epon using a standard Epon-embedding protocol including 4 h of incubation in 2% OsO_4 (EM TP, Leica). Samples were placed on a Parafilm-covered plate and covered with an Epon-filled gelatin capsule. Polymerization was performed at 60 °C for 24 h. Samples were trimmed, and semi-thin 500 nm sections were cut on a Leica UC7 ultramicrotome (Leica) equipped with a Histo diamond knife (Diatome). Semi-thin sections were stained with Azure II for 1 min to stain lipid-rich areas to confirm the ROI. Next, 60 nm ultrathin sections were cut on the same ultramicrotome equipped with Diatome diamond knife, ultra 35°. Ultrathin sections were transferred to Formvar-coated copper mesh grids (Science Services) and air dried. The sections were contrasted with Uranyless for 30 min (Electron Microscopy Sciences) and washed five times with ddH_2O. Samples were air dried before further storage. At least 10 images for each area of interest (magnification of ×4,000 for corpus callosum, ×7,000 for cortex) of ultrathin sections were acquired using a Zeiss EM900. Non-overlapping transmission EM micrographs were used to quantify the numbers of myelinated axons and *g*-ratios. Analyses of the images were done using Fiji. The number of myelinated axons was quantified using the cell counter plugin by counting all myelinated axons in the field of view of five pictures for each animal. For *g*-ratio analysis, at least 150 fibres per animal were randomly selected. The area of the axon and the entire myelinated fibre were measured and used to calculate the diameter. To determine the *g*-ratio, the axonal diameter was divided by the diameter of the myelinated fibre. In the results, the *g*-ratio is plotted against the axon calibre.

High-pressure freezing EM

Sample preparation by high-pressure freezing and freeze substitution was performed as previously described[69]. In brief, optic nerves from 6-month-old *Cnp*$^{-/-}$ mice and control WT mice were freshly dissected, immersed in 20% PVP in PBS and placed into HPF sample carriers (Wohlwend). After freezing using a HPM100 high-pressure freezer (Leica Microsystems), samples were embedded in Epon after freeze substitution using 0.1% tannic acid in acetone followed by 2% OsO_4 and 0.1% uranyl acetate in acetone. After polymerization, samples were sectioned using a UC7 ultramicrotome (Leica Microsystems) and imaged with a LEO912 transmission electron microscope (Carl Zeiss Microscopy) using an on-axis 2k CCD camera (TRS).

Cell fractionation

For analysis of APP and TREM2 fragmentation, cell fractioning was performed before western blot analysis according to published protocols[36]. In brief, tissue was homogenized in DEA-buffer (0.25% diethylamine and 50 mM NaCl pH 10) using the Precellys bead-milling method (Precellys soft tissue homogenizing lysis kit, Bertin Instruments), and the soluble protein fraction was extracted by centrifugation (10 min, 500*g*, 4 °C) followed by ultracentrifugation (1 h, 130,000*g*, 4 °C). The membrane-bound fraction was solubilized in RIPA buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 1% SDS, 2.5 mM sodium pyrophosphate and 1 mM Na_2EDTA) and cleaned by centrifugation (10 min, 500*g*, 4 °C) and ultracentrifugation (1 h, 130,000*g*, 4 °C). RIPA-insoluble material (containing plaque Aβ) was resuspended in ice-cold 70% formic acid in water, sonicated and ultracentrifuged (1 h, 130,000*g*, 4 °C). Supernatant was collected as the formic acid fraction and neutralized with 1 M Tris pH 9.5. All buffers and solutions were supplemented with protease inhibitor cocktail (P8340, Merck). Fractions were stored at −80 °C until further use.

SDS–PAGE and western blotting

To determine the protein concentration in samples, detergent-compatible protein assays (Bio-Rad) were carried out in duplicate. Samples were mixed with Laemmli sample buffer (2% SDS, 10% glycerol, 0.0025% bromphenol blue, 0.125 M Tris-Cl, pH 6.8 and 0.05 M DTT) and an equal amount of

protein (typically 20–30 µg) was loaded per lane. For BACE1 western blotting, standard Tris-glycine SDS–PAGE gels (8%) were used. For western blot analysis of APP and TREM2 fragmentation, Tris-tricine SDS–PAGE gels (10–20%, Novex, Thermo Fisher) were used. Gels were run at 100–120 V for approximately 1 h. For Tris-glycine SDS–PAGE gels, proteins were transferred onto low-fluorescent Immobilon-FL membrane (0.45 µm pore size, Merck) using the Bio-Rad wet-blot system (1.15 h, 500 mA) and blot transfer buffer (25 mM Tris and 190 mM glycine) containing 20% methanol. For Tris-tricine SDS–PAGE gels, proteins were transferred onto a low-fluorescent PVDF membrane of lower pore size (0.2 µm). Blots were washed in water, and transferred protein was stained with Fastgreen for transfer quality-check and for normalization purposes. For this, membranes were transferred to Fastgreen working solution (0.0005% Fastgreen FCF (Serva)) in de-staining solution (30% methanol, 7% ml glacial acetic acid in 63% H_2O) for 5 min and briefly washed two times in de-staining solution. Membranes were imaged using a ChemoStar fluorescent imager (Intas) equipped with a 670 nm/20 nm excitation filter and near-infrared emission collection. Membranes were rinsed in TBS with Tween (0.05%) until the pH was neutral and blocked in 5% BSA in TBS-T for 1 h at room temperature. Membranes were then incubated in primary antibodies in 5% BSA overnight at 4 °C on a rotating shaker. The following primary antibodies were used: anti-BACE1 (1:1,000, rabbit, D10E5, Cell Signaling Technologies); anti-C-terminal APP (1:1000; rabbit, A8717, Merck); anti-C-terminal APP (1:1,000; rabbit, 127-003, Synaptic Systems); anti-APP/Aβ (1:1,000; mouse, 6E10, BioLegend); anti-C-terminal TREM2 (1:1,000; rabbit, E7P8J, Cell Signaling Technologies); anti-CNP (mouse, AMAb91072, Atlas; 1:1,000); and anti-MBP (rabbit, serum, custom-made in the Nave Laboratory; 1:1,000). Membranes were washed several times, and membranes were incubated in secondary antibody solution (5% BSA in TBS-T). The following secondary antibodies were used: anti-rabbit IgG (H+L) DyLight 800 (1:1,000; Thermo Fisher) and anti-mouse IgG (H+L) Dylight 680 (1:1,000; Thermo Fisher). Membranes were scanned using an Odyssey platform (Licor). For visualization, scanned images were adjusted for brightness and contrast for optimal display. For quantification (performed on raw images), background was subtracted and bands were analysed using Fiji (integrated density). Protein levels were normalized to Fastgreen whole protein or to full-length protein in the case of APP.

MACS of microglia and bulk RNA-seq

Microglia were isolated from mouse hemibrains (excluding cerebellum and olfactory bulb) by MACS. Dissected tissues were enzymatically and mechanically dissociated using a Miltenyi Biotec adult brain dissociation kit according to manufacturer's protocol. Before microglial isolation using CD11b microglia microbeads and LS columns (Miltenyi Biotec), astrocytes (ACSA-2 microbeads) and oligodendrocytes (O4 microbeads) were removed to enhance the purity of the microglial population. Isolated microglia were directly eluted into RLT lysis buffer, and RNA was isolated using a RNeasy Micro kit (Qiagen). In total, $n = 4$ replicates were used for each genotype (WT, *Cnp*$^{-/-}$, 5×FAD and *Cnp*$^{-/-}$;5×FAD). RNA extracted from sorted mouse brain hemisphere microglia was eluted in 30 µl nuclease-free water and subjected to 50 bp single-end mRNA sequencing using HiSeq 4000 (Illumina). Raw sequencing data were first evaluated using FASTQC (v.0.72) for quality, then aligned against the reference mouse genome GRCm38 using STAR (v.2.5.2b-2)[70] with default parameters. Gene raw counts of each sample were extracted using featureCounts (v.1.6.3)[71] from aligned profiles for differential gene expression analysis using DESeq2 (v.1.26.0)[72] and converted to TPM value for sample distance calculation and visualization, as well as for gene expression pattern analysis. For differential gene expression analysis, each pair of genotype was calculated separately, and statistics results were summarized (Supplementary Table 1). Gene targets with adjusted P value < 0.05 were considered as significantly regulated. Using normalized TPM profiles, samples were embedded by PCA to assess distances.

For significantly regulated genes between each pair of genotype, functional and pathway enrichment analysis using over-representation analysis approach was performed using R (v.3.6.3) with the gprofiler2 package (v.0.2.0)[73], in which gene set enrichment analysis was performed using the WebGestalt interface[74]. DEGs obtained from all paired comparisons were aggregated and clustered based on their expression patterns across genotypes, whereby clusters were identified using k-means ($k = 10$) and expression levels were represented by scaled TPM values. GO (biological process) enrichment analysis was performed for each DEG cluster using gprofiler2, and significant terms enriched from all clusters were input for

similarity analysis using simplifyEnrichment package (v.1.5.2)[75] built under R (v4.1.0). To view potentially gene–gene networks of significantly regulated targets (adjusted $P < 0.01$) between *Cnp*$^{-/-}$;5×FAD mice and 5×FAD mice, STRING (v.11.5) analysis was applied with the highest confidence (0.900) of interaction score.

In vitro phagocytosis assay

An Aβ phagocytosis assay was performed on bone-marrow derived macrophages (BMDMs) to assess phagocytosis capabilities of phagocytes after myelin pretreatment. BMDMs were isolated and cultured as previously described[64]. For experiments, cells were detached by accutase treatment and seeded onto poly(l-lysine)-coated, HCl washed 12-mm-diameter glass coverslips (density of 18,000 cells per coverslip). Cells were left untreated (control) or treated with purified myelin membranes (purification by sucrose-gradient ultracentrifugation) at 2.5 μg ml^{-1} for 12 h. Cells were subsequently treated with HiLyte488-labelled aggregated, synthetic $A\beta_{42}$ (Anaspec, AS-60479-01) at 1 μg μl^{-1} for 4 h. Aggregation of $A\beta_{42}$ was induced by previous incubation of $A\beta_{42}$ at 100 μg ml^{-1} in DMEM at 37 °C for 24 h after vortexing the solution for 15 s. Cells were fixed in 4% PFA and PBS and subjected to fluorescence labelling with DAPI and Lectin-Dylight 649 (Lycopersicon Esculentum Tomato; Fisher Scientific, NC1093063). Lectin-Dylight 649 was incubated for 2 h at room temperature in PBS. Optical sections were acquired with a confocal laser-scanning microscope (Zeiss LSM 900 AiryScan) using a LD C-Apochromat ×40/1.1 water objective (zoom factor of 1.0). The *z*-step in *z*-stacks was kept at 0.8 μm. Images were acquired in bidirectional mode (Aryscan Mode SR:4.1 2D Auto) and the LSM scan speed was kept at 7. Laser intensity was kept at 2% 640 nm, 0.1% 561 nm, 0.5% 488 nm and 0.2% 405 nm. Image resolution was selected to be 1,000 × 1,000 pixels at 16 bit. Acquired images were maximum-intensity projected along the *z*-axis, and the amount of phagocytosed $A\beta_{42}$ was determined by thresholding in the respective channel using Fiji. This percentage positive area was normalized to the area covered by cells (as determined by lectin labelling). Two to three images per treated coverslip were analysed.

Isolation of nuclei and single-nuclei transcriptome sequencing

We generated two snRNA-seq datasets for the analysis of microglia states induced by myelin dysfunction and amyloidosis. For the first dataset, cortex and corpus callosum from 3-month-old WT, forebrain shiverer mice ($Foxg1^{cre}Mbp^{fl/fl}$), $Cnp^{-/-}$ mice and WT controls were microdissected, and tissues collected from two animals were pooled for each replicate. For the second dataset, brain hemispheres (without cerebellum and olfactory bulb) were extracted from 6-month-old WT, $Cnp^{-/-}$, 5×FAD and $Cnp^{-/-}$;5×FAD mice. Two replicates per genotype were sequenced. Nuclei were isolated according to previously published methods[76]. In brief, frozen tissue was transferred into 2 ml of pre-chilled homogenization buffer (320 mM sucrose, 0.1% NP40, 0.1 mM EDTA, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 10 mM pH 7.8 Tris, 167 µM β-mercaptoethanol and 1× protease inhibitor (Roche)). Tissue was carefully homogenized and filtered through a 80 µm strainer and further centrifuged for 1 min at 100 .r.c.f. For each sample, 400 µl supernatant was collected into a pre-chilled 2 ml low-binding Eppendorf tube, followed by adding 400 µl 50% iodixanol solution (in 1× homogenization buffer containing 480 mM sucrose) to reach a 25% iodixanol concentration. By layering 600 µl of 29% iodixanol underneath the 25% iodixanol mixture, then 600 µl of 35% iodixanol underneath the 29% iodixanol layer, two clear interfaces between different concentrations of buffers were created, and the tube was centrifuged for 20 min at 3,000 r.c.f. After centrifugation, nuclei were collected from the band between the 29% and the 35% iodixanol layers and transferred to a fresh pre-chilled tube. Isolated nuclei were washed and resuspended in cold resuspension buffer (1× PBA, 1% BSA and 0.1 U μl^{-1} RNase inhibitor) and further subjected to single-nuclei transcriptome libraries using a chromium single cell 3′ reagent kit according to the manufacturer's instructions (10x Genomics). The constructed libraries were sequenced using Novaseq 6000 (Illumina). Raw snRNA-seq data were collected in Fastq format. Data generated from the 3-month time point was aligned to reference genome premRNA using CellRanger toolkit (v.3.0.2; 10x Genomics). The 6-month time point snRNA-seq data were first aligned to the reference mm10 genome version index from 10x Genomics (refdata-gex-mm10-2020-A) using CellRanger toolkit (v.6.1.2; 10x Genomics), and aligned profiles were further processed using CellBender (v.0.2.0) for removal of ambient transcripts.

Matrices containing a unique molecular identifier (UMI) count of each gene in each nuclei were extracted for all samples by filtering out nuclei with <200 detected genes and <500 total transcripts, as well as nuclei with outlier level transcript quantity or a gene detection rate identified according to individual sample sequencing depth (Supplementary Table 1). Genes expressed in fewer than three cells were excluded from further analysis. Filtered expression matrices were combined based on data time point (3-month and 6-month), and the UMI of each nucleus were normalized towards its total UMI count with a scale factor of 10,000 and then log transformed.

Dimensionality reduction, clustering analysis and cell type annotation

The normalized UMI matrix for 3-month and 6-month time point data were mainly analysed using the R package Seurat (v.4.1.1)[77,78]. Highly variable genes were calculated and scaled to support linear dimensionality reduction using PCA. For all cells sequenced from 3-month-old animals, the first 50 PCs were used for further neighbouring embedding using UMAP[79], as well as for performing the clustering analysis with a resolution of 0.5 using *K*-nearest neighbour algorithm. Cells sequenced from 6-month-old animals underwent the same neighbouring embedding protocol using the first 50 PCs and proceeded with clustering analysis with a resolution of 0.8. Cluster marker genes in both datasets were calculated using both default Wilcoxon test and the MAST algorithm[80] to carefully determine cluster cell-type annotations. Clusters with undefined cell identities were removed from further analysis. To perform cell-type-specific analysis for microglia, the corresponding cell population from each dataset was first subset and reduced for its dimensionality using PCA. Similarly, selected top PCs were used for UMAP embedding and clustering analysis, with cluster marker genes calculated using the default Wilcoxon test. Specific parameters used for analysing microglia can be found in Supplementary Table 1.

External human snRNA-seq data microglia subset re-analysis

Previously published human snRNA-seq data—snRNAseqAD_TREM2 (ref. [35]) and snRNAseqPFC_BA10 (ref. [46])—were obtained from AD Knowledge Portal (https://adknowledgeportal.org) following application for data access. Aligned data profiles were downloaded and pre-processed with quality control and removal of outlier cells, referring to parameters that were used in the respective original publications. Cells from each dataset were analysed using Seurat (v.4.1.2). Microglia and macrophage cell populations were identified by a set of marker genes including *CX3CR1*, *P2RY12*, *TMEM119*, *HEXB*, *TREM2* and *SPP1* and further extracted for subset analysis. Microglia subset expression matrices from both datasets underwent data normalization, high variable gene calculation, linear reduction of dimensionality and eventually neighbouring embedding using UMAP. Unbiased clustering analysis was performed at a resolution of 0.5 in both subsets. Cluster marker genes were calculated using default setups compiled in the Seurat analysis pipeline. Subpopulations that were identified as doublets, low-quality cells or contaminations were removed from further analysis and visualizations.

Data visualization

Images were exported from the respective imaging or bioinformatics software (ZEN 2011 blue edition, Zeiss; Vision4D Arivis; Fiji or R) and final figures were assembled in Inkscape (v.1.1, https://www.inkscape.org). Graphs were created using Prism 8.0 (GraphPad).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw sequencing data and raw and processed count matrices have been uploaded to the Gene expression Omnibus[81] under the SuperSeries accession number GSE178304 (microglia bulk RNA-seq in GSE178296; 6-month-old snRNA-seq data in GSE208683; 3-month-old snRNA-seq data in GSE178295). External datasets recruited for analysis were accessed through accession numbers listed in the Methods. Source data are provided with this paper.

Code availability

Packages involved in data analysis pipelines are listed in the Methods. The code used for bulk and single-nuclei transcriptome sequencing is available on GitHub at https://github.com/TSun-tech/AD_MyelinMutant.git. The Fiji macro script for analysis of plaque-corralling microglia has also been deposited into GitHub at https://github.com/TSun-tech/AD_MyelinMutant.git. More detailed information is available upon contacting the corresponding authors.

References

1. Bowley, M. P., Cabral, H., Rosene, D. L. & Peters, A. Age changes in myelinated nerve fibers of the cingulate bundle and corpus callosum in the rhesus monkey. *J. Comp. Neurol.* **518**, 3046–3064 (2010).

2. Safaiyan, S. et al. Age-related myelin degradation burdens the clearance function of microglia during aging. *Nat. Neurosci.* **19**, 995–998 (2016).

3. Safaiyan, S. et al. White matter aging drives microglial diversity. *Neuron* **109**, 1100–1117.e10 (2021).

4. Griffiths, I. et al. Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science* **280**, 1610–1613 (1998).

5. Fünfschilling, U. et al. Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature* **485**, 517–521 (2012).

6. Lee, Y. et al. Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* **487**, 443–448 (2012).

7. Saab, A. S. et al. Oligodendroglial NMDA receptors regulate glucose import and axonal energy metabolism. *Neuron* **91**, 119–132 (2016).

8. Toyama, B. H. et al. Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971–982 (2013).

9. Ando, S., Tanaka, Y., Toyoda, Y. & Kon, K. Turnover of myelin lipids in aging brain. *Neurochem. Res.* **28**, 5–13 (2003).

10. Yeung, M. S. et al. Dynamics of oligodendrocyte generation and myelination in the human brain. *Cell* **159**, 766–774 (2014).

11. Ringman, J. M. et al. Diffusion tensor imaging in preclinical and presymptomatic carriers of familial Alzheimer's disease mutations. *Brain* **130**, 1767–1776 (2007).

12. Dean, D. C. et al. Association of amyloid pathology with myelin alteration in preclinical Alzheimer disease. *JAMA Neurol.* **74**, 41–49 (2017).

13. Wang, Q. et al. Quantification of white matter cellularity and damage in preclinical and early symptomatic Alzheimer's disease. *NeuroImage Clin.* **22**, 101767 (2019).

14. Araque Caballero, M. Á. et al. White matter diffusion alterations precede symptom onset in autosomal dominant Alzheimer's disease. *Brain* **141**, 3065–3080 (2018).

15. Snaidero, N. et al. Antagonistic functions of MBP and CNP establish cytosolic channels in CNS myelin. *Cell Rep.* **18**, 314–323 (2017).

16. Edgar, J. M. et al. Early ultrastructural defects of axons and axon–glia junctions in mice lacking expression of *Cnp1*. *Glia* **57**, 1815–1824 (2009).

17. Trevisiol, A. et al. Structural myelin defects are associated with low axonal ATP levels but rapid recovery from energy deprivation in a mouse model of spastic paraplegia. *PLoS Biol.* **18**, e3000943 (2020).

18. Bush, A. I. et al. Rapid induction of Alzheimer Aβ amyloid formation by zinc. *Science* **265**, 1464–1467 (1994).

19. Jankowsky, J. L. et al. Rodent Aβ modulates the solubility and distribution of amyloid deposits in transgenic mice. *J. Biol. Chem.* **282**, 22707–22720 (2007).

20. Chen, J.-F. et al. Enhancing myelin renewal reverses cognitive dysfunction in a murine model of Alzheimer's disease. *Neuron* **109**, 2292–2307.e5 (2021).

21. Zhang, X. et al. Oligodendroglial glycolytic stress triggers inflammasome activation and neuropathology in Alzheimer's disease. *Sci. Adv.* **6**, eabb8680 (2020).

22. Chung, J. A. & Cummings, J. L. Neurobehavioral and neuropsychiatric symptoms in Alzheimer's disease: characteristics and treatment. *Neurol. Clin.* **18**, 829–846 (2000).

23. Cherny, R. A. et al. Treatment with a copper–zinc chelator markedly and rapidly inhibits β-amyloid accumulation in Alzheimer's disease transgenic mice. *Neuron* **30**, 665–676 (2001).

24. Frenkel, D., Maron, R., Burt, D. S. & Weiner, H. L. Nasal vaccination with a proteosome-based adjuvant and glatiramer acetate clears β-amyloid in a mouse model of Alzheimer disease. *J. Clin. Invest.* **115**, 2423–2433 (2005).

25. Schoenemann, P. T., Sheehan, M. J. & Glotzer, L. D. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.* **8**, 242–252 (2005).

26. Stokin, G. B. et al. Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. *Science* **307**, 1282–1288 (2005).

27. Gowrishankar, S., Wu, Y. & Ferguson, S. M. Impaired JIP3-dependent axonal lysosome transport promotes amyloid plaque pathology. *J. Cell Biol.* **216**, 3291–3305 (2017).

28. Vagnoni, A. et al. Calsyntenin-1 mediates axonal transport of the amyloid precursor protein and regulates Aβ production. *Hum. Mol. Genet.* **21**, 2845–2854 (2012).

29. Niederst, E. D., Reyna, S. M. & Goldstein, L. S. Axonal amyloid precursor protein and its fragments undergo somatodendritic endocytosis and processing. *Mol. Biol. Cell* **26**, 205–217 (2015).

30. Buxbaum, J. D. et al. Alzheimer amyloid protein precursor in the rat hippocampus: transport and processing through the perforant path. *J. Neurosci.* **18**, 9629–9637 (1998).

31. Lazarov, O., Lee, M., Peterson, D. A. & Sisodia, S. S. Evidence that synaptically released β-amyloid accumulates as extracellular deposits in the hippocampus of transgenic mice. *J. Neurosci.* **22**, 9785–9793 (2002).

32. Gowrishankar, S. et al. Massive accumulation of luminal protease-deficient axonal lysosomes at Alzheimer's disease amyloid plaques. *Proc. Natl Acad. Sci USA.* **112**, E3699–E3708 (2015).

33. Sadleir, K. R. et al. Presynaptic dystrophic neurites surrounding amyloid plaques are sites of microtubule disruption, BACE1 elevation, and increased Aβ generation in Alzheimer's disease. *Acta Neuropathol.* **132**, 235–256 (2016).

34. Yuan, P. et al. *TREM2* haplodeficiency in mice and humans impairs the microglia barrier function leading to decreased amyloid compaction and severe axonal dystrophy. *Neuron* **90**, 724–739 (2016).

35. Zhou, Y. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).

36. Parhizkar, S. et al. Loss of TREM2 function increases amyloid seeding but reduces plaque-associated ApoE. *Nat. Neurosci.* **22**, 191–204 (2019).

37. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).

38. Hammond, T. R. et al. Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity* **50**, 253–271.e6 (2019).

39. Deming, Y. et al. The *MS4A* gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci. Transl. Med.* **11**, eaau2291 (2019).

40. Huynh, T.-P. V., Davis, A. A., Ulrich, J. D. & Holtzman, D. M. Apolipoprotein E and Alzheimer's disease: the influence of apolipoprotein E on amyloid-β and other amyloidogenic proteins: thematic review series: ApoE and lipid homeostasis in Alzheimer's disease. *J. Lipid Res.* **58**, 824–836 (2017).

41. Schmued, L. C., Raymick, J., Paule, M. G., Dumas, M. & Sarkar, S. Characterization of myelin pathology in the hippocampal complex of a transgenic mouse model of Alzheimer's disease. *Curr. Alzheimer Res.* **10**, 30–37 (2013).

42. Mitew, S. et al. Focal demyelination in Alzheimer's disease and transgenic mouse models. *Acta Neuropathol.* **119**, 567–577 (2010).

43. Behrendt, G. et al. Dynamic changes in myelin aberrations and oligodendrocyte generation in chronic amyloidosis in mice and men. *Glia* **61**, 273–286 (2013).

44. Chen, W.-T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991.e19 (2020).

45. Kenigsbuch, M. et al. A shared disease-associated oligodendrocyte signature among multiple CNS pathologies. *Nat. Neurosci.* **25**, 876–886 (2022).

46. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

47. Janova, H. et al. Microglia ablation alleviates myelin-associated catatonic signs in mice. *J. Clin. Invest.* **128**, 734–745 (2018).

48. Jendresen, C. et al. Systemic LPS-induced Aβ-solubilization and clearance in AβPP-transgenic mice is diminished by heparanase overexpression. *Sci. Rep.* **9**, 4600 (2019).

49. Wendeln, A.-C. et al. Innate immune memory in the brain shapes neurological disease hallmarks. *Nature* **556**, 332–338 (2018).

50. Sheng, J. G. et al. Lipopolysaccharide-induced-neuroinflammation increases intracellular accumulation of amyloid precursor protein and amyloid β peptide in APPswe transgenic mice. *Neurobiol. Dis.* **14**, 133–145 (2003).

51. Knopp, R. C., Baumann, K. K., Wilson, M. L., Banks, W. A. & Erickson, M. A. Amyloid beta pathology exacerbates weight loss and brain cytokine responses following low-dose lipopolysaccharide in aged female Tg2576 mice. *Int. J. Mol. Sci.* **23**, 2377 (2022).

52. Mahmoudi, E. et al. Diagnosis of Alzheimer's disease and related dementia among people with multiple sclerosis: large cohort study, USA. *Mult. Scler. Relat. Disord.* **57**, 103351 (2022).

53. Luczynski, P., Laule, C., Hsiung, G.-Y. R., Moore, G. W. & Tremlett, H. Coexistence of multiple sclerosis and Alzheimer's disease: a review. *Mult. Scler. Relat. Disord.* **27**, 232–238 (2019).

54. Bartzokis, G. Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. *Neurobiol. Aging* **25**, 5–18 (2004).

55. Braak, H. & Braak, E. Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. *Acta Neuropathol.* **92**, 197–201 (1996).

56. Oakley, H. et al. Intraneuronal β-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).

57. Saito, T. et al. Single *App* knock-in mouse models of Alzheimer's disease. *Nat. Neurosci.* **17**, 661–663 (2014).

58. Lappe-Siefke, C. et al. Disruption of *Cnp1* uncouples oligodendroglial functions in axonal support and myelination. *Nat. Genet.* **33**, 366–374 (2003).

59. Klugmann, M. et al. Assembly of CNS myelin in the absence of proteolipid protein. *Neuron* **18**, 59–70 (1997).

60. Lüders, K. A., Patzig, J., Simons, M., Nave, K. A. & Werner, H. B. Genetic dissection of oligodendroglial and neuronal *Plp1* function in a novel mouse model of spastic paraplegia type 2. *Glia* **65**, 1762–1776 (2017).

61. Meschkat, M. et al. White matter integrity in mice requires continuous myelin synthesis at the inner tongue. *Nat. Commun.* **13**, 1–18 (2022).

62. Gorski, J. A. et al. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the *Emx1*-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).

63. Kawaguchi, D., Sahara, S., Zembrzycki, A. & O'Leary, D. D. Generation and analysis of an improved *Foxg1*-IRES-Cre driver mouse line. *Dev. Biol.* **412**, 139–147 (2016).

64. Berghoff, S. A. et al. Microglia facilitate repair of demyelinated lesions via post-squalene sterol synthesis. *Nat. Neurosci.* **24**, 47–60 (2021).

65. Singmann, H. et al. afex: analysis of factorial experiments. R package version 0.16-1 (2016).

66. Liebmann, T. et al. Three-dimensional study of Alzheimer's disease hallmarks using the iDISCO clearing method. *Cell Rep.* **16**, 1138–1152 (2016).

67. Wirths, O. et al. N-truncated $A\beta_{4-x}$ peptides in sporadic Alzheimer's disease cases and transgenic Alzheimer mouse models. *Alzheimer Res. Ther.* **9**, 80 (2017).

68. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

69. Weil, M.-T., Ruhwedel, T., Meschkat, M., Sadowski, B. & Möbius, W. in *Oligodendrocytes: Methods and Protocols* (eds Lyons, D. A. & Kegel, L.) 343–375 (Springer, 2019).

70. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

71. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

73. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler. *F1000Res.* **9**, ELIXIR-709 (2020).

74. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).

75. Gu, Z. & Hübschmann, D. Simplify enrichment: a Bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinformatics* https://doi.org/10.1016/j.gpb.2022.04.008 (2022).

76. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

77. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

78. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

79. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018).

80. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

81. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets —update. *Nucleic Acids Res.* **41**, D991–D995 (2012).

Acknowledgements

We thank the animal caretakers and veterinarians of the animal facility at the Max Planck Institute for Experimental Medicine (MPI-NAT); A. Fahrenholz for technical assistance; H. Braak for helpful comments on the manuscript; staff at the AD Knowledge Portal for the open access of snRNA-seq data from patients with AD; and members of the Department for Neurogenetics at the MPI-NAT and the KAGS subgroup for helpful discussions and input. C.D. was supported by a Boehringer Ingelheim Fonds PhD fellowship. Work in K.-A.N.'s laboratory was supported by the Deutsche Forschungsgemeinschaft (DFG, TRR274), the Dr Myriam and

Sheldon Adelson Medical Foundation (AMRF) and an ERC Advanced Grant (MyeliNANO).

Funding

Open access funding provided by Max Planck Society.

Author information

Author notes

1. These authors contributed equally: Constanze Depp, Ting Sun

Authors and Affiliations

1. Department of Neurogenetics, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

   Constanze Depp, Ting Sun, Andrew Octavian Sasmita, Lena Spieth, Stefan A. Berghoff, Taisiia Nazarenko, Katharina Overhoff, Swati Subramanian, Torben Ruhwedel, Wiebke Möbius, Sandra Göbbels, Gesine Saher, Hauke B. Werner, Ruth Stassart & Klaus-Armin Nave

2. Institute of Neuronal Cell Biology, Technical University Munich, Munich, Germany

   Lena Spieth, Stefan A. Berghoff, Alkmini Damkou & Mikael Simons

3. German Center for Neurodegenerative Diseases, Munich, Germany

   Lena Spieth, Stefan A. Berghoff, Alkmini Damkou, Mikael Simons, Michael Willem & Christian Haass

4. Clinical Neuroscience, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

Agnes A. Steixner-Kumar, Sahab Arinrad & Hannelore Ehrenreich

5. Department of Psychiatry and Psychotherapy, University Medical Center, Georg-August University, Göttingen, Germany

Silvia Zampar, Oliver Wirths & Dilja Krueger-Burg

6. Department of German Philology, Georg-August University, Göttingen, Germany

Maik Thalmann

7. Munich Cluster of Systems Neurology (SyNergy), Munich, Germany

Mikael Simons, Michael Willem & Christian Haass

8. Department of Neurocognitive Science, Institute of Brain Science, Nagoya City University Graduate School of Medical Sciences, Nagoya, Aichi, Japan

Takashi Saito

9. Laboratory for Proteolytic Neuroscience, RIKEN Center for Brain Science, Wako, Saitama, Japan

Takashi Saito & Takaomi Saido

10. Department of Molecular Neurobiology, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

Dilja Krueger-Burg

11. Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

Riki Kawaguchi & Daniel Geschwind

12. Metabolic Biochemistry, Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians University of Munich, Munich, Germany

Michael Willem & Christian Haass

13. Paul-Flechsig-Institute of Neuropathology, University Clinic Leipzig, Leipzig, Germany

Ruth Stassart

Contributions

C.D. and K.-A.N. conceptualized and designed the study. C.D. and A.O.S. performed the microscopy and biochemical experiments. T. Sun planned and analysed the bulk RNA-seq and snRNA-seq experiments. L.S., S.A.B. and G.S. helped perform acute demyelination experiments. A.A.S.-K. helped analyse snRNA-seq experiments. S.S. provided $Emx^{cre}/Foxg1^{cre}Mbp^{fl/fl}$ mice for crossbreedings. S.G. provided critical conceptual input and was involved in generation of 5×FAD crossbreedings. W.M. performed EM analyses and generated $Emx^{cre}/Foxg1^{cre}Mbp^{fl/fl}$ mice. S.Z. and O.W. helped analyse Aβ generation in axonal swellings. S.A. and H.B.W. provided $Emx^{cre}Plp^{fl/fl}$ mice. S.A. gave critical input on the behavioural analysis of myelin mutants. T.N. and T.R. performed conventional EM experiments. K.O., S.A.B., A.D. and M.S. supervised and performed the in situ hybridization experiments. S.A.B., L.S. and M.S. performed and supervised in vitro phagocytosis assay. M.T. performed statistical analysis of behavioural data. T. Saito and T. Saido provided APP^{NLGF} mice. D.K.-B. helped perform behavioural experiments. R.K. and D.G. performed snRNA-seq. M.W. and C.H. helped perform biochemical analysis of APP processing and provided critical experimental advice. R.S. provided human autopsy material. H.E. provided conceptual input. C.D. and T. Sun created figures. C.D., T. Sun and K.-A.N. wrote the manuscript.

Corresponding authors

Correspondence to Constanze Depp or Klaus-Armin Nave.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Cortical myelin levels and gliosis in AD patients, aged wildtype mice and myelin mutants.

Continuation of Fig. 1. (**a**) Merge and single channel images showing microglia (stained with Iba1) and myelin levels (stained with PLP/CNP) as shown in Fig. 1. (**b**) Quantification of microgliosis and myelin levels in AD patients and non-AD controls as percentage area covered in the cortical areas examined. Bars represent means; dots represent individual patients/biological replicate/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**c**) Experimental setup to study the effect of myelin dysfunction on amyloid plaque load in AD mouse models. Models of genetic myelin dysfunction, acute demyelination and genetic amyelination were combined with the *5xFAD* and App^{NLGF} model of AD and effects on amyloid deposition were

investigated *in toto* by tissue clearing and light sheet microscopy (LSM). (**d**) Myelin defects drive a premature ageing phenotype in mice. Panels show microscopic images of immunolabellings against MBP (Myelin), GFAP (Astrogliosis), and Iba1 (Microgliosis) in 6-month-old wildtype, 24-month-old wildtype, 6-month-old myelin mutant (*Cnp*$^{-/-}$ and *Plp*$^{-/y}$) mice. Regions of interest (ROI) analysed are indicated on the left. Dashed lines in the upper panel mark the pial surface above the cortex (CTX). Dashed lines in lower panels outline the corpus callosum (CC). (**e**) Heatmap representation of the quantification of myelin pathology and gliosis immunostainings as shown in (c) in myelin mutants (*Cnp*$^{-/-}$ and *Plp*$^{-/y}$) and wildtype mice at 3-months (3m), 6-months (6m) and 24-months (24m). The percentage area covered was quantified in the indicated ROIs and is presented as fold-change (FC) with respect to the 3-month-old data set. Each heatmap tile represents a single animal/biological replicate/n (n = 3 per group) and is colour-coded according to the row minimum-maximum lookup table shown below. Statistical analysis and raw values are given in Supplementary Fig. 2.

Source data

Extended Data Fig. 2 Microscopic analysis of plaque changes induced by genetic dysmyelination at 3 months of age in *5xFAD* mice, 6-month old *App*^{NLGF} mice and myelin mutants on a WT background.

Extension of Fig. 2. (**a-d**) LSM analysis of 6-month-old *Plp*$^{+/y}$*5xFAD* and *Plp*$^{-/y}$*5xFAD* brains stained for Congo Red. (**a**) Representative LSM 2D single planes Inlays show closeup images of the cortex (Ctx) and alveus (Alv). Arrows indicate small amyloid deposits in the alveus. (**b**) 3D representation of hippocampal, isocortical and alveus plaques represented as coloured centroids. (**c**) 3D cropped regions of interest of brains rendered in maximum intensity modus. (**d**) 3D quantification of plaque load in the three regions of interest normalised to the respective region in control animals. Bars represent means; dots represent biological replicates/mice/n (n = 10 for control, n = 6 for mutant). Statistical analysis: two-sided, unpaired Student’s t-test (p-values are given above bars). (**e**)

Quantifications of light sheet microscopic analysis of 3-months-old *Cnp*$^{-/-}$*5xFAD* mice. Bars represent means; dots represent biological replicates/mice/n (n = 5 for control, n = 5 for KO). Statistical analysis: two-sided, unpaired Student’s t-test (p-values are given above bars). (**f**) Quantifications of light sheet microscopic analysis of 3-months-old *Plp*$^{-/y}$*5xFAD* mice. Bars represent means; dots represent biological replicates/mice/n (n = 4 for control, n = 3 for KO). Statistical analysis: two-sided, unpaired Student’s t-test (p-values are given above bars). (**g**) Light sheet microscopic single plane closeup of the alveus region of Congo red stained brains of 3-month old *5xFAD* control mice, *Cnp*$^{-/-}$*5xFAD*, and *Plp*$^{-/y}$*5xFAD* myelin mutants. Note the absence of plaques in the alveus at this age. Arrowheads indicate the alveus. CC:Corpus callosum, Alv: Alveus. Ctx: Cortex. Hip: Hippocampus. (**h**) Light sheet microscopic analysis of *Cnp*$^{-/-}$*App*^{NLGF} brains shows enhanced plaque deposition at 6-month of age when compared to *Cnp*$^{+/-}$*App*^{NLGF} controls. Left panel shows LSM single plane and a closeup of a cortical region. Right panel shows 3D maximum intensity projection of the cropped isocortical ROI. Plaque burden was quantified using machine-learning-based segmentation of amyloid plaques. Bars represent means; dots represent biological replicates/mice/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student’s t-test (p-value is given in graph). (**i**) Light sheet microscopic analysis of plaque burden in aged (14-months old) *Plp*$^{-/y}$ brains on a *WT* background. As a positive control, a 14-months old *5xFAD* brain with ample plaque pathology is shown. In aged *WT* and *Plp*$^{-/y}$ only non-specific nuclei or lipofuscin signals can be observed. No plaque-like structures could be identified in the whole brain. The same observations were made in three independent samples/mice for *WT* and *Plp*$^{-/y}$. (**j**) Immunofluorescence staining of β-sheet positive plaques (Methoxy-04) and amyloid-β (4G8) in aged (22-months old) forebrain-specific PLP knockout animals (*Emx-Cre Plp*$^{fl/fl}$). Note ample lipofuscin accumulation (autofluorescence) colocalising fully with signal in the 4G8 channel. No plaque-like structures could be identified in both groups. The same observations were made for four independent samples/mice per group.

Source data

Extended Data Fig. 3 Myelin status in AD mouse models and myelin mutant crossbreedings and their behavioural analysis.

(**a**) Representative EM images of regions of the cingulate cortex and medial corpus callosum in *App*^{NLGF} and *5xFAD* mice at 6 months of age. (**b**) Quantification of myelinated axon counts per field of view (FOV) of EM images as shown in (a). Dots represent single micrographs analysed. 5 images per animal/biological replicat/n were analysed (n = 5 for *WT* and *App*^{NLGF}, n = 6 for *5xFAD*). Contoured dots represent the mean for each biological replicate. Statistical analysis: ordinary one-way ANOVA (p-value given in graphs) with Tukey's multiple comparisons (no significant differences were found). (**c**) Analysis of myelin thickness by g-ratio measurements of single axons in the medial corpus callosum of *WT, 5xFAD* and *App*^{NLGF} mice at 6 months of age. Dots represent single axons analysed (*WT* = 477, *5xFAD* = 512, *App*^{NLGF} = 533 from 3 animal/biological replicat/n per group); lines represent linear regression, dotted line represent 95% confidence bands. In the bar graph, bars represent means and dots represent biological replicates/mice/n (n = 3). Statistical analysis: ordinary one-way ANOVA (p-value given in graphs) with Tukey's multiple comparisons. (**d**) Western blot analysis of myelin proteins MBP and CNP in *5xFAD* and *App*^{NLGF} mice at 6 months of age in different brain regions (hemisphere, cortex, hippocampus). Molecular weight marker (in kDa) is indicated on the left. (**e**) Quantification of western blot analysis shown in (d). Relative protein abundance was determined by normalisation to total protein staining and wildtype levels. Bars represent means; dots represent biological replicates/mice/n (n = 4 per group). Statistical analysis: ordinary one-way ANOVA per group (antibody, brain region). p-values are given above bar graphs. (**f**) Closeups of myelin (anti-MBP) labelling in 6-months old *WT, 5xFAD* and *App*^{NLGF} mice in (DG), cornu ammonis (CA1), corpus callosum (CC), and cortex (CTX). (**g**) Quantification of myelin profiles (MBP^+ area) in DG, CA1 and CTX and intensity in CC. Bars represent means; dots represent biological replicates/mice/n (n = 6 for *WT*, n = 6 for *5xFAD*, n = 5 for *App*^{NLGF}). Statistical analysis: ordinary one-way ANOVA with Tukey's multiple comparison test was performed. No significant ($p < 0.05$) differences were found. (**h**) Immunofluorescence closeups of myelin (MBP labelled) in plaque proximity (labelled with 6E10) in *WT, 5xFAD* and

App^{NLGF} cortex. The same observations were made in different independent samples/mice per group (n = 6 for *WT*, n = 6 for *5xFAD*, n = 5 for *App*^{NLGF}). (**i**) Closeups of myelin (anti-PLP) labelling in *Cnp*$^{-/-}$*5xFAD* mice, corresponding single mutants (*Cnp*$^{-/-}$ and *5xFAD*) and wildtype (*WT*) controls in the CA1, CTX and CC at 6 months of age. (**j**) Closeups of myelin (anti-CNP labelling in *Plp*$^{-/y}$*5xFAD* mice, corresponding single mutants (*Plp*$^{-/y}$ and *5xFAD*) and wildtype (*WT*) controls in the CA1, CTX and CC region at 6 month of age. (**k**) Quantification of myelin profiles (PLP+ area) in CA1 and CTX in *Cnp*$^{-/-}$*5xFAD* brains shown in (i). Bars represent means; dots represent biological replicates/mice/n (n = 5 for *WT*, n = 6 for *5xFAD*, n = 3 for *Cnp*$^{-/-}$, n = 5 for *Cnp*$^{-/-}$*5xFAD)*. Statistical analysis: ordinary one-way ANOVA with Tukey's multiple comparison tests. P-values for $p < 0.05$ are shown in the bar graph. Non-significant p-values are not shown. (**l**) Quantification of myelin profiles (CNP+ area) in CA1 and CTX in *Plp*$^{-/y}$*5xFAD* brains shown in (j). Bars represent means; dots represent biological replicates/mice/n (n = 4 per group). Statistical analysis: ordinary one-way ANOVA with Tukey's multiple comparison tests. P-values for $p < 0.05$ are shown in the bar graph. Non-significant p-values are not shown. (**m**) Representative tracks and images for the behaviour of *Cnp*$^{-/-}$*5xFAD* female mice in the elevated plus maze (EPM), Y maze (YM) and the clasping test paradigms. (**n**) Quantification of the behaviour of *Cnp*$^{-/-}$*5xFAD* female mice in the paradigms shown in (m). Bars represent means; dots represent biological replicates/mice/n (n = 10 for *WT*, n = 8 for *5xFAD*, n = 9 for *Cnp*$^{-/-}$, n = 10 for *Cnp*$^{-/-}$*5xFAD*). (**o**) Representative tracks and images for the behaviour of Plp$^{-/y}$5xFAD male mice in the elevated plus maze (EPM), Y maze (YM) and the clasping test paradigms. (**p**) Quantification of the behaviour of *Plp*$^{-/-}$*5xFAD* female mice in the paradigms shown in (m). Bars represent means; dots represent biological replicates/mice/n (n = 10 for *WT*, n = 8 for *5xFAD*, n = 8 for *Plp*$^{-/y}$, n = 10 for *Plp*$^{-/y}$*5xFAD*). Statistical analysis for (**n**) and (**p**): two-way type III ANOVAs probing the main effects for the *5xFAD* genotype (p_{AD}) and the respective *myelin-mutant* genotype (p_{MY}) as well as their interaction (p_{INT}) followed by Tukey's multiple comparisons posthoc test (p-value is given in the bar graphs).

Source data

Extended Data Fig. 4 Cuprizone and EAE induced changes in brain myelin status and its influence on brain plaque burden.

(**a**) Histological analysis of Cuprizone-induced brain demyelination and microglia/macrophage infiltration after 4-week Cuprizone treatment by Iba1 and PLP immunostaining. Dashed lines border indicated ROIs: latCC = lateral corpus callosum. medCC = medial corpus callosum. Alv = Alveus. Fim = Fimbria. Closeup of immunostaining showing specific demyelination and microglia/macrophage accumulation of the medial corpus callosum and alveus after 4 weeks of Cuprizone treatment. Quantification of immunostaining for PLP (myelin) and microglia/macrophage (Iba1) in the alveus as shown in (a) throughout 2, 3 and 4 weeks of Cuprizone treatment. Bars represent means; dots represent biological replicates/mice/n (n = 3 per group). Dashed line represents mean background fluorescence of PLP staining. (**b**) Histological analysis of remyelination after a 4-week recovery period after Cuprizone withdrawal by Iba1 and PLP immunostaining. Bars represent means; dots represent biological replicates/mice/n (n = 6 per group). FStatistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**c**) LSM single plane of Congo Red stained brains and closeups of the subiculum and a cortical region. Middle panel shows 3D distribution of isocortical and hippocampal plaques represented as centroids. Lower panel shows 3D maximum intensity projection of cropped regions of interest. Right lower panel shows 3D quantifications of plaque burden parameters in hippocampus and isocortex. Bars represent means; dots represent biological replicates/mice/n (n = 6 for control, n = 5 for Cup). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**d**) Neurological scoring shows successful EAE-induction in *5xFAD* animals with typical disease onset at around day 10 post induction (dpi 10). Lines represent time courses for single animals/replicates/mice/n (n = 5). Dots represent daily scores. (**e**) 3D quantifications of plaque burden in the brain of EAE *5xFAD* animals 4 weeks dpi (age 14 weeks). Bars represent means; dots represent biological replicates/mice/n (n = 4 for control, n = 5 for EAE *5xFAD*). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**f**) Immunofluorescence analysis of potential brain lesions in the EAE brain using nuclei staining (DAPI) and anti-MBP staining. No classical lesions (accumulation of nuclei and corresponding loss of myelin) were observed in the brain. Lower panel shows closeup of

the area indicated in the middle panel. (**g**) Immunofluorescence analysis of gliosis in the EAE brain using Iba1 and GFAP staining. (**h**) Quantification of myelin content and gliosis (as shown in (f) and (g)) in cortex and corpus callosum of the EAE brain. Bars represent means; dots represent biological replicates/mice/n (n = 3 for control, n = 4 for EAE). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**i**) Light sheet microscopic analysis of plaque burden in the lumbar spinal cord of EAE *5xFAD* mice. No typical grey matter plaques could be detected in 14 week old *5xFAD* or EAE *5xFAD* mice. The lumbar spinal cord was heavily affected by EAE lesions as visualised by Iba1 staining.The same observations were made in three independent samples/mice per group (n = 4 for control, n = 5 for EAE *5xFAD*). (**j**) As a positive control for successful detection of spinal cord plaques by light sheet microscopy, a cervical spinal cord of a 6-month-old *5xFAD* animal is shown.

Source data

Extended Data Fig. 5 Recombination territories and plaque load in *forebrain shiverer 5xFAD* mice.

(**a**) Basic characterisation of myelination in *Emx-Cre Mbp$^{fl/fl}$5xFAD* mice. Autofluorescence shows a clear lack of myelinated profiles in the corpus callosum (arrows) while the thalamus and striatum show normal myelin profiles. The lower panel shows closeup images of anti-CNP and MBP co-immunolabelling in *Emx-Cre Mbp$^{fl/fl}$5xFAD*. Lack of myelin compaction (MBP^{-}CNP^{+}) throughout the cortex, hippocampus and subcortical white matter. Compaction of myelin is unaffected in other brain regions such as the thalamus and striatum (MBP^{+}CNP^{+}). The same observations were made in 5 independent samples/mice per group. (**b-d**) *In toto plaque* load in 3-month-old forebrain *shiverer* 5xFAD (*Emx-Cre Mbp$^{fl/fl}$ 5xFAD*). (**b**) Representative LSM 2D single planes show closeup images of the cortex (Ctx) and alveus (Alv). (**c**) 3D representation of hippocampal and isocortical represented as coloured centroids. (**d**) 3D cropped regions of interest of brains rendered in maximum intensity modus and 3D quantification of plaque load in the three regions of interest normalised to the respective region in control animals. Bars represent means; dots

represent biological replicates/mice/n (n = 9 for control, n = 9 for mutant). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**e**) Quantification of light sheet microscopic analysis of plaque load in 6-month-old *Emx-Cre* $Mbp^{fl/fl}$*5xFAD* mice. Bars represent means; dots represent biological replicates/mice/n (n = 4 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**f**) Immunofluorescence labelling of amyloid plaques in the white matter of 6-month old *5xFAD* control mice and different myelin mutant *5xFAD* crossbreedings. Note the differences of white matter plaques in the amyelination model (*Emx-Cre* $Mbp^{fl/fl}$*5xFAD)* versus dysmyelination models *(*$Cnp^{-/-}$/$Plp^{-/y}$ *5xFAD)* in regards to their distribution and appearance: While in $Cnp^{-/-}$ and $Plp^{-/y}$ crossbreedings enhanced plaque deposition is restricted to the alveus (two lower dashed lines), in *Emx-Cre* $Mbp^{fl/f}$ *5xFAD* mice enhanced plaque deposition can also be seen in the corpus callosum (upper two dashed lines). Plaques in the alveus of $Cnp^{-/-}$ and $Plp^{-/y}$ crossbreedings appeared smaller and more numerous (white arrowheads), while alveus plaques in *Emx-Cre* $Mbp^{fl/fl}$ appeared bigger in size and more isolated (black contoured arrow heads). cc:corpus callosum, alv: alveus. The same observations were made in three independent samples/mice per group.

Source data

Extended Data Fig. 6 Myelin dysfunction alters APP processing in $Cnp^{-/-}$*5xFAD* and $Plp^{-/y}$ *5xFAD* mice.

Continuation from Fig. 3. (**a**) Overview of APP-stained swellings in the fimbria of *5xFAD* control and $Cnp^{-/-}$*5xFAD* mice. Hip: Hippocampus. Fim: Fimbria. The same observations were made in 5 independent samples/mice per group. (**b**) Colocalisation of APP and NF200 (axonal cytoskeletal marker) in swellings confirming their axonal origin in $Cnp^{-/-}$*5xFAD* mice. White arrowheads indicate swellings double-positive for APP and NF200. Black-contured arrows indicate swellings that are solely positive for APP. This is likely a result of cytoskeletal breakdown in axonal swellings. The same observations were made in 3 independent samples/mice per group. (**c**) Distinction of plaque-associated axonal swellings induced by the *5xFAD*

genotype and myelin-damage-associated swellings induced by the *Cnp*$^{-/-}$ genotype. Upper panels are closeups of images shown in Fig. 3c. Lower panels show a mask indicating axonal swellings. Note that plaque-associated swellings form a circular arrangement of multiple swellings (corona) around a plaque centre (asteriks). In contrast, myelin damage-associated axonal swellings occur isolated/scattered. (**d**) Schematic representation of plaque-associated axonal swellings and myelin damage-associated swellings, and typical stainings observed with antibodies shown in Fig. 3b. D3E10, 80C2 and 6C3 do not show cross-reactivity to full length APP and typically only very weakly stain APP-positive plaque-associated swellings, but robustly stain swellings in *Cnp*$^{-/-}$*5xFAD* mice. (**e**) Left panels shows positive staining controls for APP processing machinery antibodies used in Fig. 3 in the *5xFAD* grey matter. For APP, BACE1 and PSEN2 typical plaque-corona staining (contoured arrowheads) can be observed. For APP and BACE1 additional neuronal cell body staining (n) can be observed. Right panel shows positive staining controls for amyloid antibodies used in Fig. 3 in the *5xFAD* grey matter. Contoured arrowheads indicate proper amyloid plaques, typically stained very intensely. 029-2 and 6E10 antibodies show cross-reactivity to full-length APP and also stain plaque-associated axonal swellings (white arrowheads) and neurons (n). (**f**) 6E10 and sAPPβ_{swe} (soluble cleavage product after β-cleavage of APP) labelling of axonal swellings in *Cnp*$^{-/-}$*5xFAD* fimbria. Figure 4d continued. Quantifications are shown on the right. Bars represent means; dots represent biological replicates/mice (for sAPP$\beta_{swe,}$ n = 4 for each group; for 6E10 labelling, n = 3 for *5xFAD* controls and n = 5 for *Cnp*$^{-/-}$*5xFAD*). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graphs). (**g**) Colocalisation of BACE1 and APP/Aβ in axonal swellings in *Cnp*$^{-/-}$*5xFAD* white matter/fimbria. APP/Aβ was visualised using the 6E10 antibody. White arrowheads indicate swellings that are double-positive (to various degrees) for BACE1 and 6E10. Contoured black arrowheads indicate swellings single-positive for either 6E10 or BACE1. The same observations were made in 3 independent samples/mice per group. (**h**) sAPPβ_{swe} staining in cortical tissue of *Cnp*$^{-/-}$*5xFAD* mice and *5xFAD* controls. Staining reveals abundant plaque-associated swellings arranged in a swelling corona in 5xFAD mice. In *Cnp*$^{-/-}$*5xFAD* both plaque-associated swellings and numerous plaque-independent swellings (i.e. myelin-damage-

associated swellings) can be found. Quantification of sAPPβ$_{swe}$ covered area is shown on the right. Bars represent means; dots represent biological replicates/mice/n (n = 4 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**i**) Immunostaining analysis of APP/Amyloid-β and BACE1 in axonal swellings in *Plp*$^{−/y}$*5xFAD* white matter (fimbria). (**j**) Quantifications of immunostainings shown in (i). Bars represent means; dots represent biological replicates/mice/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**k**) Colocalisation of BACE1 and APP/Aβ in axonal swellings in *Plp*$^{−/y}$*5xFAD* mice. White arrows indicate a number of swellings in which colocalisation occurs. Contoured black arrows indicate swellings without colocalisation. The same observations were made in 3 independent samples/mice per group. (**l**) Fluorescent immunoblot analysis of BACE1 levels in the micro-dissected white matter of *Cnp*$^{−/−}$*5xFAD* and *5xFAD* mice. Molecular weight marker (in kDa) is indicated on the left. Bars represent means; dots represent biological replicates/lanes/mice/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). Source data are given in Supplementary Fig. 1. (**m**) Immunoblot analysis of APP processing using the 6E10 antibody. Molecular weight marker (in kDa) is indicated on the left. Anti-c-terminal APP labelling was applied to identify different APP CTF fragments: (p)C99 6E10 and A8717 double-positive; C89 and C83 6E10 negative and A8717 positive. Bars represent means; dots represent biological replicates/lanes/mice/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). Source data are given in Supplementary Fig. 3. (**n**) Fluorescent immunoblot analysis of APP fragmentation in the membrane-bound fraction of micro-dissected white matter (corpus callosum + alveus) of *Cnp*$^{−/−}$ *5xFAD* and *5xFAD* control mice. Molecular weight marker (in kDa) is indicated on the left. (**o**) Quantification of western blot analysis shown in (n). Bars represent means; dots represent biological replicates/lanes/mice/n (n = 3 per group). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). Source data are given in Supplementary Fig. 1.

Source data

Extended Data Fig. 7 *In vitro* and *in vivo* analysis of microglia challenged by myelin and amyloid.

(**a**) Scheme illustrating the experimental setup to study amyloid phagocytosis after myelin digestion. Bone marrow-derived macrophages (BMDMs) were isolated and treated with myelin debris for 12h (1) prior to incubation with HiLyte488-labelled $A\beta_{42}$ for 4h (2). The internalisation of $A\beta_{42}$ was assessed by microscopy. (**b**) Representative immunofluorescence images of BMDMs treated with HiLyte488-labelled $A\beta_{42}$ with and without (control) prior myelin treatment. Quantification of amyloid internalisation normalised to $Lectin^{+}$ area is given on the right. Bars represent means; dots represent replicates/coverslips/n. ($n = 3$ per group, with 2-3 quantified images per coverslips). Statistical analysis: two-sided, unpaired Student's t-test (p-value is given in graph). (**c**) Representative images of microglia immunostainings (Iba1 and DAPI) in 6-months old *$Cnp^{-/-}$5xFAD* mice and respective controls (*WT, $Cnp^{-/-}$* and *5xFAD*). Dotted line highlights the border between cortical (CTX) and callosal tissue (CC). (**d**) Quantification of microgliosis (number of microglia, $Iba1^{+}$ area) in 6-months old *$Cnp^{-/-}$5xFAD* mice and respective controls (*WT, $Cnp^{-/-}$* and *5xFAD*) in the corpus callosum (CC) and cortex (CTX). Bars represent means; dots represent biological replicates/mice ($n = 5$ for *WT*, $n = 6$ for *5xFAD*, $n = 3$ for *$Cnp^{-/-}$*, $n = 5$ for *$Cnp^{-/-}$5xFAD*) Statistical analysis: ordinary one-way ANOVA with Tukey's posthoc test for multiple comparisons (p-values for the posthoc test for all $p < 0.05$ are given in the graphs). (**e**) Representative images of microglia immunostainings (Iba1 and DAPI) in 6-months old *$Plp^{-/y}$5xFAD* mice and respective controls (*WT*, *$Plp^{-/y}$* and *5xFAD*). Dotted line highlights the border between cortical (CTX) and callosal tissue (CC). (**f**) Quantification of microgliosis (number of microglia, $Iba1^{+}$ area) in 6-months old *$Plp^{-/y}$5xFAD* mice and respective controls (*WT*, *$Plp^{-/y}$* and *5xFAD*) in the corpus callosum (CC) and cortex (CTX). Bars represent means; dots represent biological replicates/mice ($n = 4$ for each experimental group). Statistical analysis: ordinary one-way ANOVA with Tukey's posthoc test for multiple comparisons (p-values for the posthoc test for all $p < 0.1$ are given in the graphs). (**g**) Microscopic analysis of microglial corralling by amyloid (6E10), microglia (Iba1) and nucleus

(DAPI) co-labelling in 6-month-old *Cnp*$^{-/-}$*5xFAD* and *5xFAD* control mice. Upper panel: overview images of the Iba1 staining. Lower panel: closeup of a cortical region. Quantification of microglia corralling (number of microglia per plaque) in the cortex is given on the right. Violin plots show the distributions for single biological replicates/mice/n (n = 3 per genotype, 50 plaques per replicate). Black lines in violin plots indicate medians. Bar graph on the right shows the comparison of means (n = 3 per genotype, 50 plaques per replicate). Statistical analysis: Two-sided, unpaired Student's t-test (p-value is given above graph). Right, lower panel: closeup of a representative microglia-plaque interaction in each genotype. (**h**) Microscopic analysis of microglial corralling by amyloid (6E10), microglia (Iba1) and nucleus (DAPI) co-labelling in 6-month-old *Cnp*$^{-/-}$*App*^{NLGF} and *Cnp*$^{+/-}$*App*^{NLGF} control mice. Upper panel: overview images of the Iba1 staining. Lower panel: closeup of a cortical region. Quantification of microglia corralling (number of microglia per plaque) in the cortex of *Cnp*$^{-/-}$*App*^{NLGF} and *Cnp*$^{+/-}$*App*^{NLGF}. Violin plots show the distributions for single biological replicates/mice/n (n = 3 per genotype, 50 plaques per replicate). Black lines in violin plots indicate medians. Bar graph on the right shows the comparison of means (n = 3 per genotype, 50 plaques per replicate). Line indicates means. Statistical analysis: Two-sided, unpaired Student's t-test (p-value is given above graph). Right, lower panel: closeup of a representative microglia-plaque interaction per genotype. (**i**) Microscopic analysis of microglial corralling by amyloid (6E10), microglia (Iba1) and nucleus (DAPI) co-labelling in 6-month-old *Plp*$^{-/y}$*5xFAD* and *5xFAD* control mice. Upper panel: overview images of the Iba1 staining. Lower panel: closeup of a cortical region. (**j**) Quantification of microglia corralling (number of microglia per plaque) in the cortex of 6-month old *Plp*$^{-/y}$*5xFAD* and *5xFAD*. Violin plots show the distributions for single biological replicates/mice/n (n = 3 per genotype, 50 plaques per replicate). Black line in violin plots indicate medians. Bar graph on the right shows the comparison of means (n = 3 per genotype, 50 plaques per replicate). Line indicates mean. Statistical analysis: Two-sided, unpaired Student's t-test (p-value is given above graph). (**k**) Quantifications of microglia corralling (microglia coverage per plaque) in the cortex of 6-month-old *Plp*$^{-/y}$*5xFAD* and *5xFAD*. Violin plots show the distributions for single biological replicates/mice/n (n = 3 per genotype; 2010 plaques for *5xFAD* and 2529

plaques for *Plp*$^{-/y}$*5xFAD* genotype in total). Black line in violin plots indicate medians. Bar graph on the right shows the comparison of medians (n = 3 per genotype). Line indicates mean. Statistical analysis: Two-sided, unpaired Student's t-test (p-value is given above graph). (**l**) Microscopic analysis of microglial corralling in Cuprizone-treated *5xFAD* animals by amyloid (6E10), microglia (Iba1) and nucleus (DAPI) co-labelling. Note residual white matter gliosis in Cuprizone *5xFAD* mice and small amyloid deposits highlighted by white arrowheads. The same observations were made in different independent samples/mice per group (n = 4 for *5xFAD* and n = 5 for Cuprizone *5xFAD)*. (**m**) Microscopic analysis of microglial corralling in EAE *5xFAD* animals by amyloid (6E10), microglia (Iba1) and nucleus (DAPI) co-labelling. A comparison of a non-lesion site versus a lesion site is shown. Note the massive microglia/macrophage infiltration at the lesion site that made corralling analysis difficult. The same observations were made in different independent samples/mice per group (n = 5 for *5xFAD* and n = 5 for Cuprizone *5xFAD)*.

Source data

Extended Data Fig. 8 Analysis of microglia bulk RNA sequencing data from *Cnp*$^{-/-}$ *5xFAD*, single mutants and *WT* controls.

Extension of Fig. 4. (**a**) Venn diagram presentation of shared/unique differentially expressed genes (DEGs) between experimental groups (compared to *WT*). (**b**) Myelin-related transcript cluster in the top100 gene list contributing to principal component 1 (PC1). Note that the detection of myelin-related transcripts most likely results from myelin contamination in the CD11b-MACS purified fraction. Decreased contamination (i.e. higher purity) in *5xFAD*, *Cnp*$^{-/-}$ and *Cnp*$^{-/-}$*5xFAD* is presumably linked to the increased number of microglia cells and advanced tissue disintegration in inflammatory conditions. Each heatmap tile represents a single animal/biological replicate/n (n = 4 per genotype). (**c**) Volcano plot representation of differentially expressed genes in single mutants (*5xFAD*, *Cnp*$^{-/-}$) and double mutant *Cnp*$^{-/-}$*5xFAD* in comparison to *WT* animals (significant cutoff adjP < 0.01, log2FC > 1.0), (**d**) Line plots for the 10 DEG

groups identified by k-means clustering (based on scaled TPM values) with unique trajectories throughout genotypes/experimental groups. Lines represent connections between individual genotype data. Error bars represent SD. (**e**) Heatmap of normalised mean expression pattern in the 10 DEG clusters across different genotypes. Each cluster of genes underwent GO enrichment analysis using gprofiler2, and significantly enriched terms were analysed for their similarities. The left-side coloured heatmap indicates GO term significant levels across each cluster. The adjacent grayscale heatmap shows cluster similarities based on their enriched GO terms. Most representative keywords shared by enriched GO terms are given in the text boxes on the right side. (**f**) Volcano plot representation of differentially expressed genes between *Cnp*$^{-/-}$*5xFAD* and *5xFAD* microglia (significant cutoff adjP < 0.01, log2FC > 1.0). (**g**) Top100 differentially regulated genes (based on significant levels) between *Cnp*$^{-/-}$*5xFAD* and *5xFAD* microglia. Left panel shows genes upregulated in *Cnp*$^{-/-}$*5xFAD*. Right panel shows genes downregulated in *Cnp*$^{-/-}$*5xFAD*. (**h**) Correlation analysis of DEGs in *Cnp*$^{-/-}$*5xFAD* and *5xFAD* in comparison to *WT*, respectively, reveals highly correlated gene regulations. Pearson correlation coefficient (R) and corresponding p-value is given in the graph. (**i**) Functional enrichment analysis of DEGs from *Cnp*$^{-/-}$*5xFAD* vs *5xFAD* using gprofiler and gene set enrichment analysis (GSEA). For each analysis, gene ontology biological process and pathway enrichment is given, and the top 20 significant terms are presented as barplots. The colour indicates the proportion of genes involved in each enriched term. (**j**) STRING interaction analysis of DEGs (significant cutoff adjP < 0.01) between *Cnp*$^{-/-}$*5xFAD* and *5xFAD* (confidence 0.9, n = 219 nodes with connections are displayed). The colours of the edges indicate interaction form.

Source data

Extended Data Fig. 9 snRNA-seq analysis of *Cnp*$^{-/-}$ *5xFAD* and respective single mutant controls.

Continuation of Fig. 5. (**a**) Left panel shows Western blot analysis of Trem2 levels in *WT vs 5xFAD* mice demonstrating clear induction of full length

Trem2 (30–60 kDa, glycosylated forms) and enhanced Trem2 cleavage (~12 kDa) in *5xFAD* mice. Lanes present single animals/biological replicates/n (n = 3 per group). The right panel shows Western blot analysis of Trem2 levels in *Cnp*$^{-/-}$ *5xFAD* and *5xFAD* mice. No differences in Trem2 cleavage could be detected. Total protein fastgreen staining is shown as loading control. Lanes present single animals/biological replicates/n (n = 3 per group). Source data are given in Supplementary Fig. 3. (**b**) Cell type distribution and annotations from 6-month-old mouse brain hemisphere snRNA-seq data visualised in the UMAP space. In total 4 genotypes (*WT, Cnp*$^{-/-}$, *5xFAD*, and *Cnp*$^{-/-}$*5xFAD*) were included and 61,949 cells passed quality controls for downstream analysis. Right panel shows microglial cells highlighted in yellow in UMAP space. (**c**) Heatmap visualisation of scaled marker gene expressions at single-cell level. For each cell population, a random n = 1000 cells were selected to be displayed. For NFOL and pericyte, all cells are shown (total cell numbers from both populations were under 1000). (**d**) Violin plots showing gene expressions across oligodendroglia and microglia subpopulations (microglia detailed subpopulations annotated in Fig. 5b). (**e**) *Ms4a7* fluorescent single molecule in situ hybridisation (RNAscope) in combination with Iba1 immunostaining on *5xFAD* and *Cnp*$^{-/-}$*5xFAD* brain tissue. The same observations were made in different independent samples/mice per group (n = 3 per group). (**f**) 2D zoom in on UMAP representation of amyloid and myelin DAM colour coded to cluster annotation (left panel) or genotype (right panel). (**g**) 3D representation of DAM clusters colour-coded according to genotype. (**h**) Volcano plot representation of DEGs between DAM subclusters and between genotypes within Amyloid- and Myelin-DAM clusters. (**i**) Heatmap representation of genes commonly regulated in-between DAM clusters and between genotypes in specific DAM clusters. (**j**) Immunofluorescence analysis of myelin phagocytosis (PLP^{+}) by microglia (Iba1^{+}) in the cortex of *Cnp*$^{-/-}$*5xFAD* mice. Quantification on the right shows the number of myelin-positive(PLP^{+}) microglia (Iba1^{+}). Bars represent means, dots represent single mice/biological replicates/n (n = 3 for *5xFAD*, n = 4 for *Cnp*$^{-/-}$*5xFAD*). Statistical analysis: Two-sided, unpaired Student’s t-test (p-value is given in the graph).

Source data

Extended Data Fig. 10 Lack of microglia activation and APP metabolism changes in *forebrain shiverer 5xFAD* mice and MS4A cluster gene expression in human AD microglia.

(**a**) Experimental setup to study microglia states in *FB-shiv* (*forebrain shiverer)* in comparison to wildtype controls and *Cnp*$^{-/-}$ brain by snRNAseq at 3 months of age. (**b**) Identification of microglia/macrophage subpopulations and genotype distribution visualised in the UMAP space. (**c**) Marker gene expression for the four identified clusters visualised in violin plots. (**d**) Number of microglia per microglia subtype and genotype. Different bars represent single animals/biological replicates (n = 2 per group). (**e**) Assessment of microglia coverage and morphology by Iba1 stainings in 3-month-old *FB shiverer 5xFAD* mice. Lower panel shows closeups of cells delineated by dashed lines in the upper panel. As a comparison and positive control for activated microglia a micrograph of *Cnp*$^{-/-}$*5xFAD* is shown on the right. Bargraphs show quantification of Iba1^{+} area and microglia number in the cortex (CTX) and corpus callosum (CC). Bars represent means, dots represent animals/biological repolicates/n (n = 4 for control and n = 3 for *FB shiverer* mice). Statistical analysis: two-sided Student's t-test (p-value is given in the graph). (**f**) Assessment of APP+ axonal swellings in the fimbria (FIM) of 3-month-old *FB shiverer 5xFAD* by 4G8 (APP/Aβ) immunolabeling. In both control *5xFAD* mice and *FB shiverer 5xFAD* APP+ swellings were rarely or not found. The contoured arrow in the left micrograph highlights an amyloid plaque. As a comparison and positive control for APP+ swellings a micrograph of *Cnp*$^{-/-}$*5xFAD* is shown on the right. Swellings are highlighted by arrowheads. Bars represent means, dots represent animals/biological repolicates/n (n = 4 for control and n = 3 for *FB shiverer* mice). Statistical analysis: two-sided Student's t-test (p-value is given in the graph). (**g**) Western blot analysis of APP processing in *FB shiverer 5xFAD* mice. Bands for full-length APP (fAPP) and c-terminal fragments (CTFs) are shown. n = 3 biological replicates/mice/lanes per group. Source data are given in Supplementary Fig. 1. Molecular weight marker (in kDa) is given on the left. (**h**-**k**) Reanalysis of microglia subcluster of Mathys et al.[46]. (**h**) Genotype and clustering distribution of the microglia/leukocyte subset in the Mathys et al.[46] snRNA-seq dataset visualised in the UMAP space. Hom = homeostatic

microglia, DAM = Disease-associated microglia, Ribo high = high in ribosomal transcripts. (**i**) Histogram representation of cell numbers in each subcluster according to sample condition (control versus AD). Bars represent cell numbers. (**j**) Expression analysis of classical DAM marker genes (*TREM2, SPP1, APOE*) and *MS4A6A* in the main microglial subcluster (leukocytes not shown) visualised in the UMAP space (left) and by dot plot representation (right). (**k**) Co-expression analysis of MS4A6A and SPP1 at the single-cell level visualised in the UMAP space. Colour indicates gene expression and co-expression level according to the colour scale shown on the right. (**l-o**) Reanalysis of microglia subcluster of and Zhou et al.[35]. (**l**) Genotype and clustering distribution of the microglia/leukocyte subset in the Zhou et al.[35] snRNA-seq dataset visualised in the UMAP space. Hom = homeostatic microglia, DAM = Disease-associated microglia, Histone high = high in histone transcript, MyTE = Myelin transcript enriched. (**m**) Histogram representation of cell numbers in each subcluster according to sample condition (control versus AD). (**n**) Expression analysis of classical DAM marker genes (*TREM2, APOE*), *ABCA1*, and *MS4A6A* in the main microglial subcluster (leukocytes, monocytes and MyTE not shown) visualised in the UMAP space (left) and by dot plot representation (right). (**o**) Co-expression analysis of *MS4A7* and *TREM2* at the single-cell level visualised in the UMAP space. Colour indicates gene expression and co-expression level according to the colour scale shown on the right. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org).

Source data

Supplementary information

Supplementary Information

Supplementary Guide and Supplementary Figs. 1–3.

Reporting Summary

Supplementary Table 1

Supplementary Table 2

Source data

Source Data Figs. 2–5 and Source Data Extended Data Figs. 1–10.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Depp, C., Sun, T., Sasmita, A.O. *et al.* Myelin dysfunction drives amyloid-β deposition in models of Alzheimer's disease. *Nature* **618**, 349–357 (2023). https://doi.org/10.1038/s41586-023-06120-6

- Received: 30 July 2021

- Accepted: 21 April 2023
- Published: 31 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06120-6

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

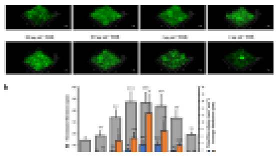
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

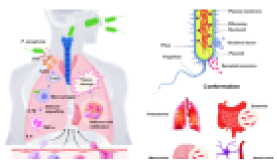
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06120-6



Extracellular DNA release, quorum sensing, and PrrF1/F2 small RNAs are key players in Pseudomonas aeruginosa tobramycin-enhanced biofilm formation

23 May 2019

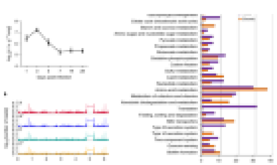
Ali Tahrioui, Rachel Duchesne, … Sylvie Chevalier



Pseudomonas aeruginosa: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics

25 June 2022

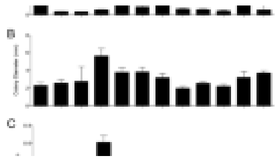
Shugang Qin, Wen Xiao, … Min Wu



TesG is a type I secretion effector of Pseudomonas aeruginosa that suppresses the host immune response during chronic infection

07 January 2019

Kelei Zhao, Wujiao Li, … Xikun Zhou



[Blocking RpoN reduces virulence of Pseudomonas aeruginosa isolated from cystic fibrosis patients and increases antibiotic sensitivity in a laboratory strain]

30 April 2019

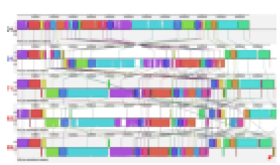
M. G. Lloyd, J. L. Vossler, … J. F. Moffat



[Evolutionary trade-offs associated with loss of PmrB function in host-adapted Pseudomonas aeruginosa]

06 July 2018

Laura Bricio-Moreno, Victoria H. Sheridan, … Daniel R. Neill



[Virulence Characteristics and an Action Mode of Antibiotic Resistance in Multidrug-Resistant Pseudomonas aeruginosa]

24 January 2019

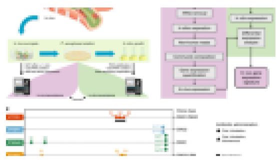
Wontae Hwang & Sang Sun Yoon



Pseudomonas aeruginosa PA80 is a cystic fibrosis isolate deficient in RhlRI quorum sensing

11 March 2021

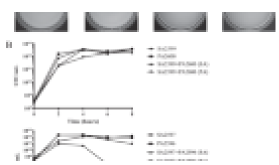
Syed A. K. Shifat Ahmed, Michelle Rudden, … James S. G. Dooley



High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs

27 August 2018

Elio Rossi, Marilena Falcone, … Helle Krogh Johansen



Coexistence with Pseudomonas aeruginosa alters Staphylococcus aureus transcriptome, antibiotic resistance and internalization into epithelial cells

12 November 2019

Paul Briaud, Laura Camus, … Karen Moreau

- Article

- Open Access
- Published: 24 May 2023

A *Pseudomonas aeruginosa* small RNA regulates chronic and acute infection

- Pengbo Cao,
- Derek Fleming,
- Dina A. Moustafa,
- Stephen K. Dolan,
- Kayla H. Szymanik,
- Whitni K. Redman,
- Anayancy Ramos,
- Frances L. Diggle,
- Christopher S. Sullivan,
- Joanna B. Goldberg,
- Kendra P. Rumbaugh &
- …
- Marvin Whiteley

Show authors

Nature **volume 618**, pages 358–364 (2023)

- 9531 Accesses
- 120 Altmetric
- Metrics details

Abstract

The ability to switch between different lifestyles allows bacterial pathogens to thrive in diverse ecological niches[1,2]. However, a molecular understanding of their lifestyle changes within the human host is lacking. Here, by directly examining bacterial gene expression in human-derived samples, we discover a gene that orchestrates the transition between chronic and acute infection in the opportunistic pathogen *Pseudomonas aeruginosa*. The expression level of this gene, here named *sicX*, is the highest of the *P. aeruginosa* genes expressed in human chronic wound and cystic fibrosis infections, but it is expressed at extremely low levels during standard laboratory growth. We show that *sicX* encodes a small RNA that is strongly induced by low-oxygen conditions and post-transcriptionally regulates anaerobic ubiquinone biosynthesis. Deletion of *sicX* causes *P. aeruginosa* to switch from a chronic to an acute lifestyle in multiple mammalian models of infection. Notably, *sicX* is also a biomarker for this chronic-to-acute transition, as it is the most downregulated gene when a chronic infection is dispersed to cause acute septicaemia. This work solves a decades-old question regarding the molecular basis underlying the chronic-to-acute switch in *P. aeruginosa* and suggests oxygen as a primary environmental driver of acute lethality.

Main

Many pathogenic bacteria can colonize their hosts and persist chronically. In some cases, bacteria disseminate from the initial infection sites to other body parts, resulting in acute systemic diseases. A central question in biology is to understand the molecular mechanisms and environmental cues controlling this lifestyle switch. The opportunistic human pathogen *P. aeruginosa* can cause both acute and chronic infections that are notoriously difficult to treat. *P. aeruginosa* exists as single cells (planktonic lifestyle) or matrix-enclosed aggregates (biofilm lifestyle), which are considered to favour acute and chronic infections, respectively. Therefore, laboratory studies have traditionally focused on probing the biofilm–planktonic transition of *P. aeruginosa* in vitro, with the goal to understand this pathogen's lifestyle changes in humans. Decades of work have shown how global regulatory systems such as the second messenger cyclic di-GMP[3,4,5] and the Gac–Rsm system[2,6,7,8,9,10,11,12] control the biofilm–planktonic transition of *P. aeruginosa* in vitro, providing critical insights into the life history of this

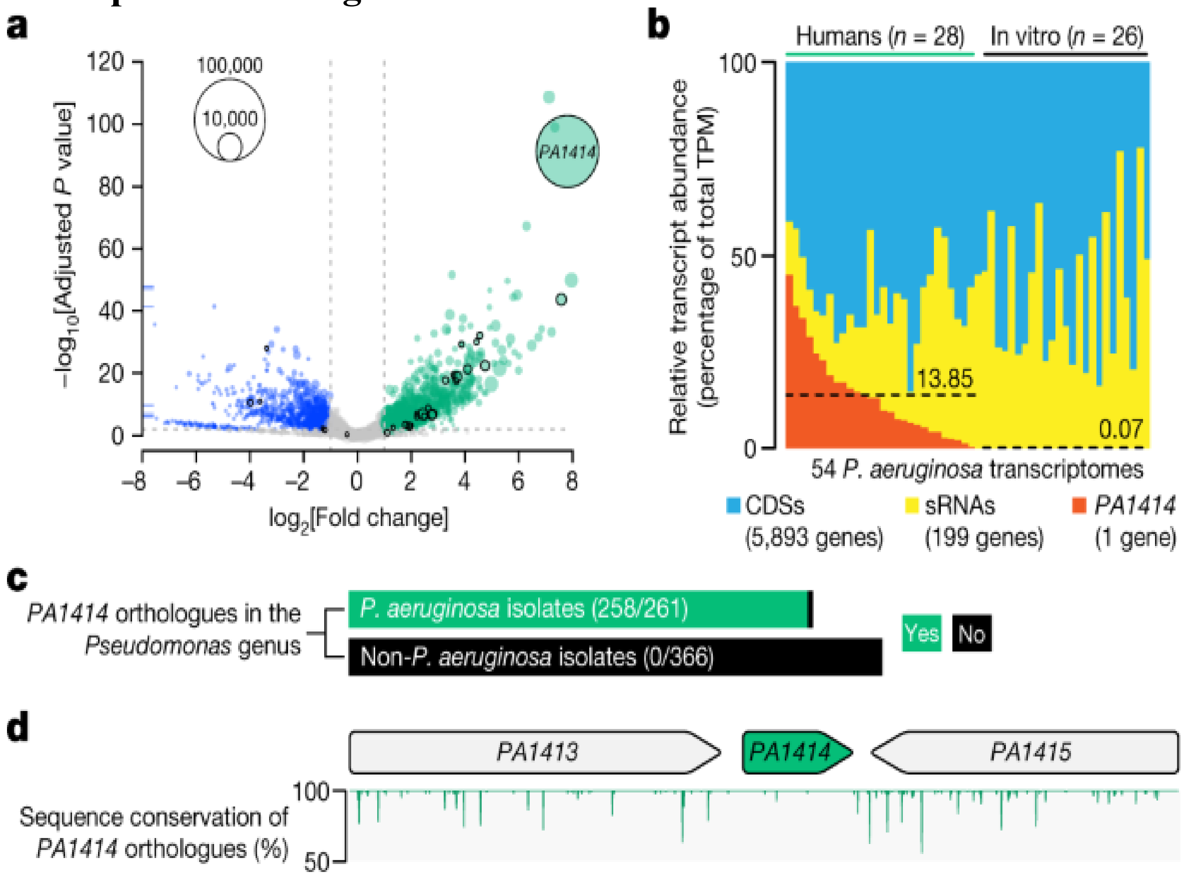
bacterium. However, so far, it has remained elusive how *P. aeruginosa* responds to environmental cues and accordingly switches infection lifestyles within the mammalian host. In this study, we addressed this gap in knowledge by leveraging *P. aeruginosa* transcriptomes acquired from human-derived samples to discover and mechanistically characterize a new small RNA, termed sRNA inducer of chronic infection X (SicX), that governs the *P. aeruginosa* chronic or acute decision during mammalian infection.

High expression of *PA1414* in humans

We previously obtained high-resolution *P. aeruginosa* transcriptomes from human infection samples, including sputum samples from patients with cystic fibrosis and debridement samples from patients with chronic wounds[13,14]. Using machine learning methods, we identified 30 *P. aeruginosa* genes whose expression levels collectively differentiate *P. aeruginosa* growth in humans from that in the laboratory[13]. More than half of these genes are uncharacterized, highlighting a substantial gap in knowledge of *P. aeruginosa* human infection biology. Included in these genes of unknown function is *PA1414* (locus tag in *P. aeruginosa* strain PAO1), whose expression level in humans was 222-fold higher than that in the laboratory, and its expression level is the highest of the *P. aeruginosa* genes expressed during human infection (Fig. 1a). We next compared the relative abundance (transcripts per million (TPM)) of *PA1414* transcripts with those of other protein-coding (5,893 genes) and non-coding (199 sRNAs)[15] transcripts (Fig. 1b). On average, *PA1414* transcripts represented 13.85% of the TPM in *P. aeruginosa* transcriptomes acquired from human chronic infection samples, and notably, they constituted almost 50% of the total TPM in some cases. However, the level of *PA1414* expression was extremely low in *P. aeruginosa* grown under standard in vitro conditions. *PA1414* is a small gene (234 base pairs long) of unknown function. Examination of 261 complete *P. aeruginosa* genomes identified 258 *PA1414* orthologues (Fig. 1c and Extended Data Fig. 1), and no homologues were identified in other *Pseudomonas* species or other organisms. The DNA sequences of *PA1414* orthologues including their upstream promoter regions are highly conserved (Fig. 1d), suggesting that *PA1414* has a conserved

function and the regulation of its expression is universal across *P. aeruginosa* isolates.

Fig. 1: The expression level of *PA1414* is the highest of the *P. aeruginosa* genes expressed during human chronic infections.



a, Differential gene expression of *P. aeruginosa* between human infections and common in vitro conditions. For each gene, the $\log_2$[fold change] is plotted against the Wald test *P* value. Bubbles with black borders highlight the 30 signature genes of *P. aeruginosa* human infection. Bubble size indicates mRNA read abundance during human infections. Blue bars on the *y* axis indicate genes (with negative $\log_2$[fold change] values) outside the scale range. Dashed grey lines indicate the cutoffs ($-\log_{10}$[adjusted *P* value] > 2, $|\log_2$[fold change]$| > 1$) for identifying differentially expressed genes. The upregulated (green) and downregulated (blue) genes in humans are colour-coded differently. Grey bubbles indicate genes that are not differentially expressed. **b**, Relative transcript abundance (TPM) of *PA1414* compared to those of all other protein-coding sequences (CDSs) and non-

coding sRNAs in 54 transcriptomes examined in **a**, sorted from the highest to the lowest *PA1414* TPM. Average *PA1414* TPM is indicated with dashed lines. **c**, *PA1414* orthologues are found only in *P. aeruginosa*. Green and black bars indicate the presence and absence of orthologues, respectively. **d**, DNA sequence conservation of *PA1414* and its neighbouring genes among 258 *PA1414*-containing *P. aeruginosa* isolates. Below, the percentage of orthologues identical to the *PA1414* allele from PA14 at each nucleotide is shown.

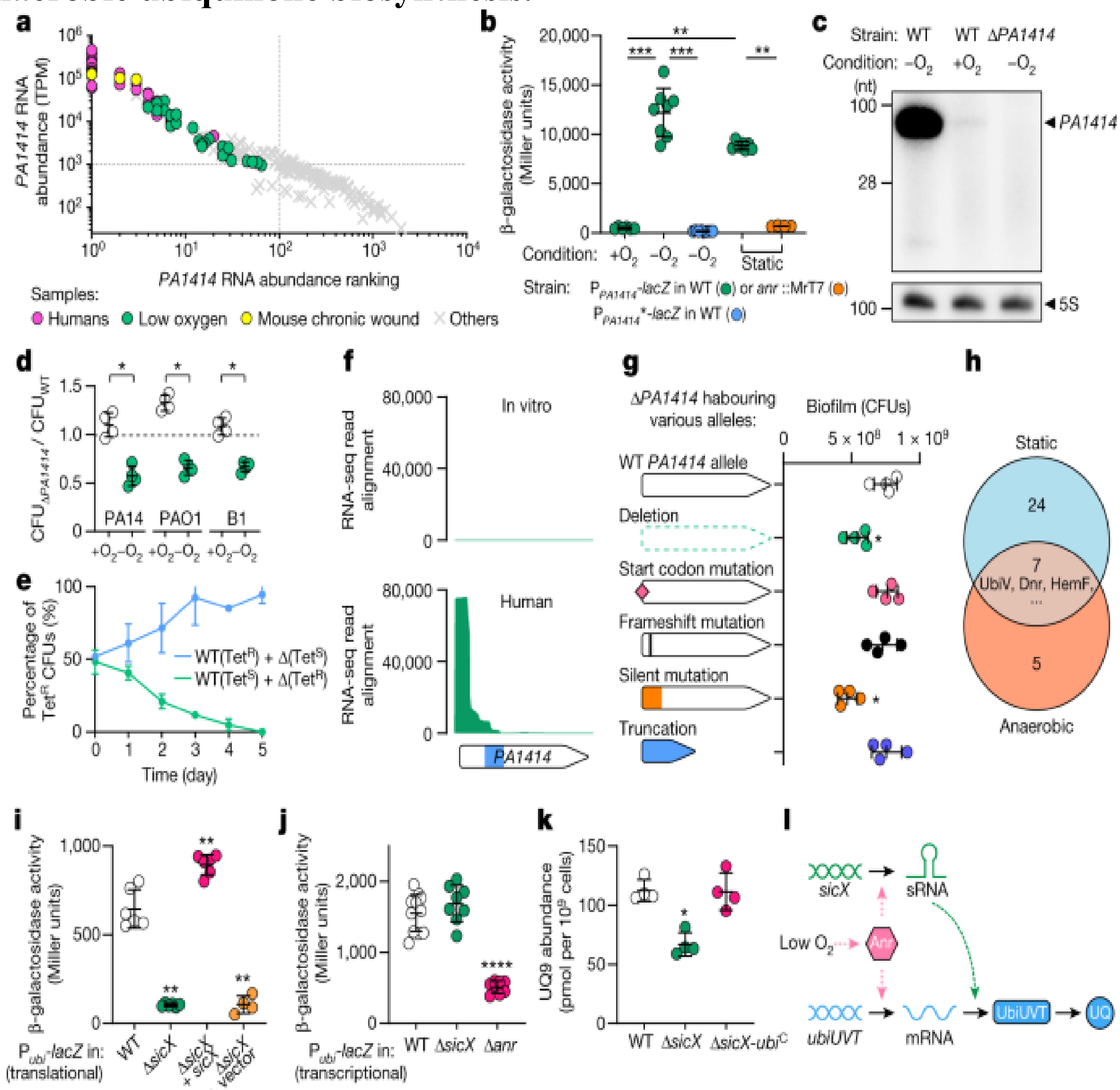
Source Data

Low oxygen induces *PA1414* expression

P. aeruginosa can thrive in diverse ecological niches within the human body. To understand what environmental cues drive *PA1414* expression, we examined 202 publicly available *P. aeruginosa* transcriptomes acquired across a range of environments (Fig. 2a and Supplementary Table 1). We found that *PA1414* was expressed at high levels under low-oxygen and anaerobic conditions. Consistent with this finding, an early microarray study also showed that *PA1414* is induced during anaerobic growth[16]. Moreover, a consensus binding motif for the global anaerobic transcriptional regulator, Anr[16], was identified upstream of *PA1414* (Extended Data Fig. 2a). Using a *lacZ* reporter that quantifies transcription of *PA1414*, we showed that indeed this gene was induced by 25-fold under anaerobic conditions, and mutagenesis of Anr or the binding motif abolished *PA1414* expression (Fig. 2b). *PA1414* was also expressed at an approximately 18-fold higher level in static cultivation (Fig. 2b), indicating that strict anaerobic conditions are not required for its induction. In addition, northern blot analysis provided direct evidence that *PA1414* transcripts were highly abundant during anaerobic but not aerobic growth (Fig. 2c). To directly compare *PA1414* expression levels in humans and in low-oxygen environments, we conducted RNA-sequencing (RNA-seq) experiments on *P. aeruginosa* during in vitro static growth (a low-oxygen condition), following the same library preparation method used in our study of *P. aeruginosa* transcriptomes in humans[14]. We found that although under low-oxygen conditions *PA1414* transcript abundance (TPM) was sevenfold lower than that during human infections, *PA1414* is among the three genes expressed at the highest levels (out of approximately 6,000

protein-coding and sRNA genes), supporting the notion that a low oxygen level is indeed a primary driver for its high expression level (Extended Data Fig. 2b). Given that *P. aeruginosa* often encounters low-oxygen environments during human chronic infection[17,18], these data indicate that *P. aeruginosa* responds to this host environmental cue by increasing the expression level of *PA1414*.

Fig. 2: *PA1414* encodes an oxygen-responsive sRNA that governs anaerobic ubiquinone biosynthesis.



a, *PA1414* RNA abundance (TPM) and the corresponding ranking (out of all protein-coding genes and sRNAs) in 202 *P. aeruginosa*

transcriptomes. Dashed lines indicate the cutoffs (*PA1414* TPM > 10^3, *PA1414* ranking < 10^2) for identifying transcriptomes with high *PA1414* expression. **b**, β-galactosidase assay examining the induction of *PA1414* under anaerobic and static conditions. P_{PA1414}-*lacZ*, *lacZ* transcriptionally fused to *PA1414* promoter; P_{PA1414}*, *PA1414* promoter containing mutations in the Anr-binding motif; MrT7, MAR2xT7 transposon. $n \geq 4$ independent experiments. **c**, Northern blot analysis of *PA1414* expression. Estimated size (nucleotides, nt) are indicated on the left. 5S rRNA served as a loading control. For gel source data, see Supplementary Fig. 1. **d**, Colony biofilm growth of Δ*PA1414* and WT in different *P. aeruginosa* strain backgrounds. The presence and absence of oxygen are indicated. The CFU of Δ*PA1414* was normalized against the CFU of the WT in each experiment ($n = 4$). A dashed line highlights the point where the CFU ratio = 1. **e**, Percentage of tetracycline-resistant (Tet^R) cells before and after daily passages under anaerobic conditions ($n = 3$). Tet^S, tetracycline sensitive; Δ, Δ*PA1414*. **f**, RNA-seq reads aligned to the *PA1414* locus during standard in vitro growth and human infections. Blue shade indicates Rho-independent terminator in *PA1414*. **g**, Colony biofilm growth of Δ*sicX* harbouring different *sicX* mutations under anaerobic conditions. $n = 4$ independent experiments. **h**, Comparative proteomic study identifies the targets of SicX under both static and anaerobic conditions. **i**, β-galactosidase assay evaluating the translational control of SicX on *ubiUVT* under anaerobic conditions ($n \geq 4$). **j**, β-galactosidase assay evaluating the transcription of *ubiUVT* in the absence of SicX or Anr under static conditions ($n \geq 8$). **k**, Quantification of anaerobic UQ9 synthesis in different strains ($n = 4$). **l**, A model of dual regulation of *ubiUVT* expression by Anr and SicX. Error bars represent standard deviation from the mean. Significant differences (compared to the WT) are indicated with asterisks (**P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; two-tailed Mann–Whitney test).

Source Data

We next considered whether *PA1414* affects the anaerobic physiology of *P. aeruginosa*. For these experiments, we deleted *PA1414* from the *P. aeruginosa* strain PA14 (the corresponding locus tag in this strain is *PA14_46160*) and compared the growth of Δ*PA1414* to that of wild-type (WT) *P. aeruginosa* under aerobic and anaerobic conditions (Fig. 2d). Using

a colony biofilm assay, we found that whereas deletion of *PA1414* did not negatively affect the aerobic growth of *P. aeruginosa*, Δ*PA1414* exhibited growth yields approximately half that of WT under anaerobic conditions with nitrate as the alternative electron acceptor (Fig. 2d). Similar observations were made in different *P. aeruginosa* strain backgrounds including a cystic fibrosis clinical isolate, B1 (Fig. 2d; ref. [19]). Although we showed that *PA1414* also played a minor role in anaerobic planktonic growth (Extended Data Fig. 2c), competition experiments in which WT and Δ*PA1414* were co-cultured planktonically (at a 1:1 starting ratio) and passaged daily under anaerobic conditions revealed that WT outcompeted Δ*PA1414* after day 3 (Fig. 2e).

PA1414 encodes an sRNA, SicX

We next investigated the molecular mechanism by which *PA1414* controls anaerobic growth. Although annotated as a protein-coding gene, we discovered that during human infection *PA1414* transcripts were restricted to the 5′ end (Fig. 2f), and a similar pattern was also observed under anaerobic conditions (Extended Data Fig. 2d). Clear enrichment of short transcripts encompassing the annotated start codon raised the question of whether *PA1414* encodes a functional protein or a small regulatory RNA (sRNA). To address this question, we introduced different mutations in *PA1414* and assessed whether they could restore the growth yield defect of the Δ*PA1414* colony biofilm under anaerobic conditions (Fig. 2g and Extended Data Fig. 2e). We found that a start codon or frame-shift mutation (to prevent protein synthesis) did not affect the function of *PA1414*, whereas synonymous substitutions in the highly transcribed region (to disrupt potential interactions of sRNA with target mRNAs or RNA-binding proteins) abolished its function. Moreover, a truncated allele encompassing only the highly transcribed 5′ region was sufficient to restore the anaerobic growth defect in Δ*PA1414*, further highlighting this functional region of *PA1414*. Through northern blot analyses, we showed that the observed growth differences were not due to differential sRNA abundance (Extended Data Fig. 2f). Thus, we conclude that *PA1414* functions as an sRNA (hereafter referred to as SicX) that is not essential for but promotes *P. aeruginosa* anaerobic growth in vitro.

SicX regulates anaerobic respiration

As sRNAs often exert their regulation post-transcriptionally[20,21,22], we carried out quantitative shotgun proteomics to identify targets of SicX. Here, proteomes of WT *P. aeruginosa* and Δ*sicX* were obtained under two growth conditions in which *sicX* is expressed at a high level: low-oxygen (that is, static growth) and strict anaerobic conditions. We identified seven proteins (including three with known functions) exhibiting different abundances between WT and Δ*sicX* under both growth conditions (Fig. 2h and Supplementary Table 2). One of these proteins is UbiV, which is encoded by the *ubiUVT* operon[23]. This operon encodes proteins responsible for the anaerobic biosynthesis of ubiquinone, an essential electron carrier for respiratory metabolism in proteobacteria[23]. Although the other two proteins encoded by this operon, UbiU and UbiT, were not in the seven proteins identified with our stringent criteria, further investigation revealed that both of these proteins also showed significantly reduced levels in Δ*sicX* compared to WT *P. aeruginosa* during static growth (Supplementary Table 2). Using a *lacZ* translational reporter of the *ubiUVT* operon, we showed that *sicX* deletion indeed reduced the translational activity of *ubiUVT* by sixfold under anaerobic conditions, and genetic complementation with *sicX* in *trans* restored *ubiUVT* translation (Fig. 2i). Moreover, SicX did not affect the transcription of *ubiUVT* (Fig. 2j). To provide direct evidence of how SicX regulates anaerobic respiration, we quantified the ubiquinone (more specifically, UQ9) abundance in WT and Δ*sicX* under anaerobic conditions by liquid chromatography–mass spectrometry, and indeed, Δ*sicX* exhibited a twofold reduction in UQ9 compared to WT (Fig. 2k). Therefore, we concluded that SicX post-transcriptionally induces the expression of UbiUVT. Our finding also provides insights into the increased abundance of Dnr and HemF in Δ*sicX* (Fig. 2h), as Dnr can induce the denitrification pathway[16] to alleviate the anaerobic stress in Δ*sicX*, and the expression of HemF is Dnr dependent[24]. Notably, we found that the transcription of *ubiUVT* is under the control of Anr (Fig. 2j). Together, these results support a model in which Anr activates the transcription of *sicX* and *ubiUVT* in response to oxygen deprivation. In turn, SicX post-transcriptionally induces ubiquinone biosynthesis to shape the anaerobic physiology of *P. aeruginosa* (Fig. 2l).

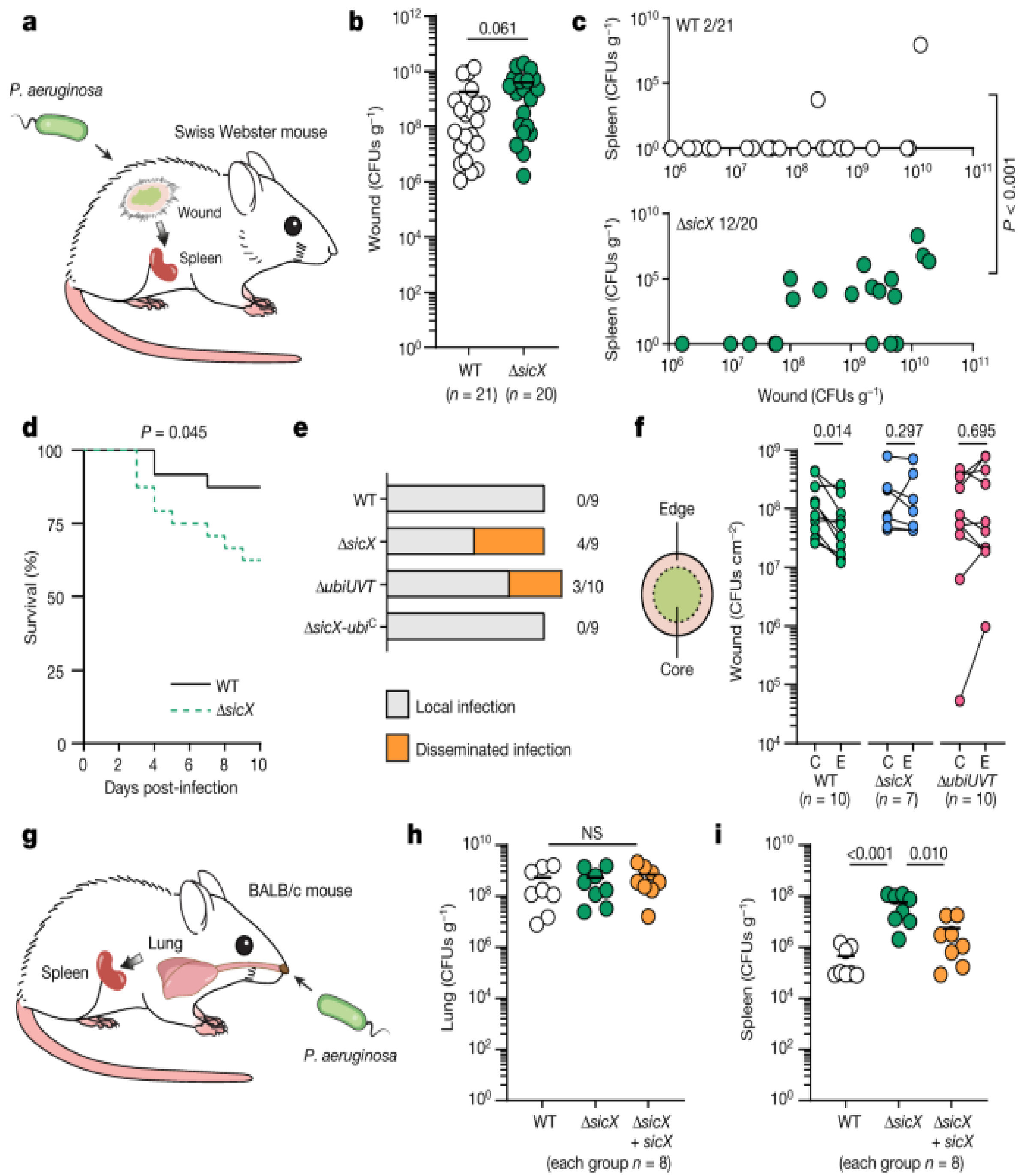
We next explored how SicX activates the translation of *ubiUVT* mRNA. In silico analysis through IntaRNA[25] identified a putative base-pairing interaction between SicX and the 5′ untranslated region (UTR) of the *ubiUVT* operon (Extended Data Fig. 3a). Secondary structures predicted with Mfold[26] indicate that a stem–loop within the 5′ UTR of *ubiUVT* occludes the ribosome-binding site of *ubiU*, whereas base pairing with SicX has the potential to liberate the ribosome-binding site for translation initiation. To experimentally test this, we carried out extensive mutational analyses of *sicX* and the 5′ UTR of *ubiUVT*, and the functional outcome of each point mutation (37 in total) was assessed through a *lacZ* reporter translationally fused to *ubiUVT* (Extended Data Fig. 3a,b). In concordance with in silico prediction, we identified mutations within the SicX seed region (for base pairing with the *ubiUVT* mRNA) and the target region in the 5′ UTR of *ubiUVT* that led to reduced translational activation of *ubiUVT*, whereas mutations in other regions generally had limited effects (Extended Data Fig. 3a,b). Moreover, a point mutation in the seed region of SicX that reduced *ubiUVT* translation can be partially restored by re-establishing base pairing in the 5′ UTR of *ubiUVT* (Extended Data Fig. 3b). However, it should be noted that the observed translational restoration may partly result from the mutation in the 5′ UTR of *ubiUVT*, which increased *ubiUVT* translation even in the absence of SicX (Extended Data Fig. 3c), probably by destabilizing the stem–loop. Finally, we showed that the RNA chaperone Hfq[27] is not required for the induction of *ubiUVT* translation by SicX (Extended Data Fig. 3d). Although Hfq is known to facilitate the interactions between sRNAs and target mRNAs in many cases, Hfq-independent sRNA–mRNA interactions have been well documented[28,29,30]. On the basis of our genetic results, we propose an Hfq-independent base-pairing model for SicX stimulation of *ubiUVT* translation, which will be tested in the future. Notably, the predicted base-pairing regions in SicX and the 5′ UTR of *ubiUVT* are identical across all examined *P. aeruginosa* genomes (Extended Data Fig. 4a–c), and thus this mechanism of regulation may be universal in *P. aeruginosa*.

SicX governs the chronic-to-acute switch

Although we now have a more granular understanding of SicX function in vitro, how SicX affects *P. aeruginosa* pathogenesis remains unknown. Here

we addressed this question by introducing WT *P. aeruginosa* and Δ*sicX* into a mouse chronic wound model (Fig. 3a). This model was chosen because it has been quantitatively shown to accurately recapitulate *P. aeruginosa* gene expression in human chronic infections[13], and SicX sRNA in this model is expressed similarly to that in humans (Fig. 2a and Extended Data Fig. 5a,b). The model involves infection of surgically created full-thickness dorsal wounds[31] (Fig. 3a). It is considered a chronic, non-lethal model because *P. aeruginosa* generally persists at the infection site for weeks without causing systemic dissemination. At 10 days post-infection, we assessed the bacterial burden in wound tissues as well as dissemination to the spleen. In wound tissues, the bacterial levels were similar between WT *P. aeruginosa* and Δ*sicX* (Fig. 3b), indicating that lacking SicX does not affect colonization. However, whereas only 2 of 21 WT-infected mice showed dissemination to the spleen, 12 of 20 Δ*sicX*-infected mice showed systemic dissemination (two-tailed Fisher's exact test, $P = 0.0009$; Fig. 3c). In addition, Δ*sicX* caused higher lethality (Fig. 3d), which is not commonly observed with WT *P. aeruginosa* in this model. Dissemination generally occurred when the wound burden was high (>10^8 colony-forming units (CFUs) g^{-1}; Fig. 3c). However, among mice carrying a high wound burden (>10^8 CFUs g^{-1}), Δ*sicX* caused more systemic infection (12 out of 15) compared to WT (2 out of 11), indicating that dissemination did not solely depend on the wound burden (two-tailed Fisher's exact test, $P = 0.0043$; Fig. 3c). Moreover, these observations are not due to differential fitness of WT and Δ*sicX* in spleens. This was demonstrated using an acute skin infection model in which systemic infection occurs shortly after subcutaneous injection into the mouse inner thigh. In this model, WT and Δ*sicX* exhibited similar colonization and growth in spleens (Extended Data Fig. 6). Taken together, our data indicate that SicX is a key player in *P. aeruginosa* chronic infection, the high expression of which promotes chronic localized infection.

Fig. 3: SicX governs chronic–acute lifestyle transition in the mammalian host.



a, Schematic of the mouse chronic wound model (10-day infection). **b**, *P. aeruginosa* burden in wounds. Lines indicate means. **c**, *P. aeruginosa* burden in the spleen (*y* axis). Corresponding wound burden is indicated on the *x* axis. **d**, Survival curves of mice infected with WT ($n = 24$) and Δ*sicX* ($n = 24$). **e**, Restoration of *ubiUVT* translation in Δ*sicX* prevented dissemination, and deletion of *ubiUVT* in WT caused dissemination. **f**, *P. aeruginosa* spatial organization in wounds. C, core; E, edge. **g**, Schematic of the mouse

pneumonia model (2-day infection). **h**, *P. aeruginosa* burden in lungs. Lines indicate means. **i**, *P. aeruginosa* burden in the spleen. Lines indicate means. Two-tailed Mann–Whitney test was used to compare the bacterial burden differences in **b**,**c**,**h**,**i**. Log-rank test and two-tailed Wilcoxon matched-pairs signed-rank test were used for **d**,**f**, respectively. Summary of $n = 6$ independent experiments for **b**–**d**; $n = 3$ independent experiments for **e**,**f**; $n = 2$ independent experiments for **h**,**i**. NS, not significant.

Source Data

To determine whether the increased dissemination and lethality of Δ*sicX* infection is linked to reduced anaerobic ubiquinone production, we first decoupled the translation of *ubiUVT* from SicX by creating a *P. aeruginosa* strain (Δ*sicX*-*ubi*^{C}) that constitutively translates UbiUVT, even in the absence of SicX (Extended Data Fig. 7a,b). This was accomplished by introducing point mutations to the 5′ UTR of *ubiUVT* that probably alter its secondary structure and enable active translation. Δ*sicX*-*ubi*^{C} grew to similar yields to those of WT *P. aeruginosa* and significantly higher yields than those of Δ*sicX* under anaerobic conditions in vitro (Extended Data Fig. 7c), and this was due to the restoration of anaerobic ubiquinone synthesis (Fig. 2k). Next, using the mouse chronic wound model, we found that Δ*sicX*-*ubi*^{C} did not exhibit systemic dissemination (Fig. 3e). We also deleted the *ubiUVT* operon from WT *P. aeruginosa*, and the resulting strain did not grow anaerobically (Extended Data Fig. 7d) owing to the defect in anaerobic ubiquinone synthesis[23]. Notably, although all strains exhibited similar bacterial burdens in the wound (Extended Data Fig. 5c), systemic dissemination was observed only in Δ*sicX* and Δ*ubiUVT* infection (Fig. 3e), indicating that sufficient anaerobic ubiquinone synthesis contributes to chronic localized infection. Although the mechanism of how ubiquinone abundance influences infection lifestyles is unclear, we found that it affects the macroscale spatial distribution of *P. aeruginosa* cells in the wound: WT cells preferentially remained in the wound centre where the cells were originally inoculated, and Δ*sicX*-infected wounds contained equivalent cells in the centre and outer edges of the wounds (Fig. 3f). This is supported by our proteomic findings showing that Δ*sicX* exhibited increased levels of two proteins (encoded by *PA0223* and *PA0224*) that are biomarkers of dispersed *P. aeruginosa*[32] (Supplementary Table 2). Moreover, Δ*ubiUVT* showed a

similar spatial organization pattern to that of Δ*sicX* (Fig. 3f). Considering that the wound probably harbours oxygen gradients, we speculate that these different spatial distributions partly reflect varying oxygen availability across the wound. Collectively, these findings suggest that reduced anaerobic ubiquinone biosynthesis promotes the dissemination of Δ*sicX*.

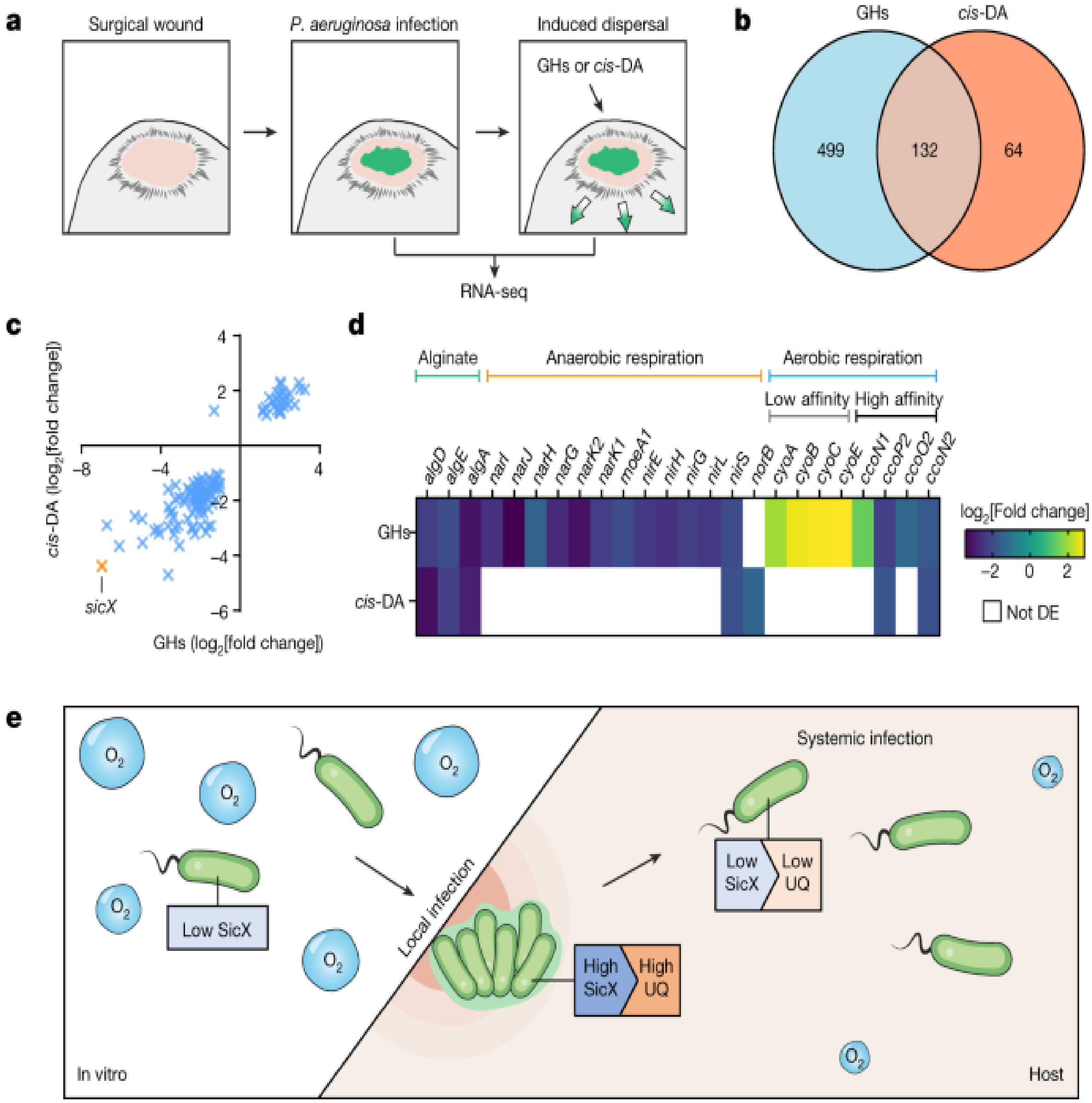
P. aeruginosa is notorious for causing a range of lung infections, and therefore, we also assessed the role of SicX in colonization and dissemination using a mouse pneumonia model (Fig. 3g). Here, *P. aeruginosa* establishes infection in the lung after intranasal administration, which in turn can lead to systemic dissemination depending on the virulence of individual strains. We found that mice infected with sublethal doses of WT and Δ*sicX* variants of PA14, a highly virulent strain, had similar bacterial burdens in the lungs at 48 h post-infection (Fig. 3h). By contrast, bacterial burdens in the spleens of Δ*sicX*-infected mice were approximately 100-fold higher than that of WT-infected mice, and complementation with the truncated *sicX* allele (depicted in Fig. 2g) in *trans* reduced dissemination (Fig. 3i), indicating that lacking SicX promotes systemic infection. Moreover, we carried out experiments using another strain, PAO1, that is less acutely virulent than PA14. We found that WT and Δ*sicX* exhibited similar colonization in the lungs, and whereas only two of nine WT-infected mice showed dissemination to the spleen, eight of nine Δ*sicX*-infected mice showed systemic dissemination (two-tailed Fisher's exact test, $P = 0.015$; Extended Data Fig. 8). These observations indicate that SicX promotes chronic localized infection in multiple mammalian infection models.

SicX is a biomarker of chronic–acute transition

As the presence or absence of SicX dictates distinct infection lifestyles, we next considered whether SicX sRNA abundance changes during the transition between chronic and acute infections. To test this, we used the mouse chronic wound model in which WT *P. aeruginosa* was first allowed to establish local infection and then treated with glycoside hydrolases (GHs) or *cis*-2-decenoic acid (*cis*-DA) to induce systemic dissemination, mimicking the chronic-to-acute transition (Fig. 4a). GHs and *cis*-DA were used as they offer distinct mechanisms that drive biofilm dispersal: GHs target biofilm exopolysaccharides by hydrolysing the glycosidic linkages to

passively induce dispersal[33,34,35], whereas *cis*-DA is a fatty acid signalling molecule that actively drives dispersal by regulating bacterial gene expression and physiology[36,37,38]. Importantly, it has been shown that GH-induced biofilm dispersal in this model can cause lethal septicaemia[35]. Here we compared *P. aeruginosa* transcriptomes before (that is, mature biofilms in wound tissues) and after (that is, dispersed cells) the treatment with GHs or *cis*-DA (Fig. 4b). Functional enrichment analyses support the notion that GHs and *cis*-DA induced different physiological and metabolic changes during in vivo dispersal (Extended Data Fig. 9). Strikingly, among the genes that were differentially expressed in response to both treatments, we found that *sicX* was the top downregulated gene (Fig. 4c), supporting the notion that SicX levels change markedly during the chronic-to-acute transition. We also found that genes involved in alginate synthesis (*algA* and *algD*) and export (*algE*) were downregulated during biofilm dispersal (Fig. 4d). Moreover, the denitrification pathway (*nir*, *nar* and molybdopterin cofactor genes in response to GHs; *nir* and *nor* genes in response to *cis*-DA) and high-oxygen-affinity oxidases (*cbb*$_3$-type oxidase in response to both GHs and *cis*-DA) were downregulated, whereas low-oxygen-affinity oxidases (Cyo oxidase in response to GHs) were upregulated (Fig. 4d), indicating an increase of oxygen availability during dispersal. Collectively, our observations identify *sicX* as a biomarker for chronic-to-acute transition and suggest oxygen as a key environmental signal during this process.

Fig. 4: SicX is a biomarker for chronic–acute lifestyle transition.



a, Schematics of in vivo dispersal of mature biofilms in mouse chronic wounds. **b**, Venn diagram of genes differentially expressed before ($n = 3$ animals) and after GH ($n = 2$) or *cis*-DA ($n = 2$) treatment. **c**, *sicX* (orange) is the top downregulated gene among the 132 differentially expressed genes identified under both GH and *cis*-DA treatments. **d**, Heat map of differentially expressed (DE) genes involved in alginate synthesis and secretion, anaerobic respiration and aerobic respiration. Aerobic terminal oxidases with low or high oxygen affinities are indicated. Blank cells indicate genes not identified as differentially expressed. **e**, A working model of how SicX allows *P. aeruginosa* to establish chronic local infection in a

host environment with low oxygen. No or low SicX promotes systemic infection.

Source Data

In conclusion, by leveraging bacterial transcriptomes from human infections, we discovered that *P. aeruginosa* produces an oxygen-responsive sRNA, SicX, the abundance of which orchestrates the transition between chronic and acute infections (Fig. 4e). The ability to respond to stressful environments is crucial for the survival of many organisms, during which sRNAs play key roles in fine-tuning gene expression. For example, in *Salmonella* infection, the sRNA PinT is induced and regulates gene expression to facilitate the transition from the invasion stage to intracellular survival[39]. In another case, *Helicobacter pylori* silences the expression of the sRNA HPnc4160 during infection, which upregulates target genes to promote bacterial colonization in the host and potentially increase the risk of gastric carcinogenesis[40]. In this study, we provide a compelling case of an sRNA directing important lifestyle choices during *P. aeruginosa* infection. Moreover, SicX shows promise as a prognostic and diagnostic biomarker, as it is the most highly responsive gene during the chronic-to-acute transition.

Although decades of in vitro studies have defined key regulatory systems governing the biofilm–planktonic transition in *P. aeruginosa*, robust evidence substantiating their roles in the mammalian host is lacking. For example, the Gac–Rsm signalling cascade is widespread in many bacterial species, and it is known to modulate the expression of features associated with planktonic or biofilm lifestyle in *P. aeruginosa*. This cascade relies on the interactions between global post-transcriptional regulators (RsmA or RsmN (also known as RsmF)) and Rsm sRNAs to fine-tune gene expression[6] (Extended Data Fig. 10a). However, we found that Rsm sRNAs are expressed only at low levels in patients chronically infected with *P. aeruginosa* (Extended Data Fig. 10b,c), indicating that they are minimally involved in maintaining long-term infection. Next, although studies have shown that inactivation of specific components of Gac–Rsm led to fitness or colonization defects in acute infection models[7,41], compelling evidence supporting a lifestyle switch model during mammalian infection remains lacking. Moreover, we found that the regulation of SicX expression is

independent from the GacS–GacA two-component system (Extended Data Fig. 10d). In addition, the primary *P. aeruginosa* strain used in this study (PA14) has a mutation in *ladS*, a sensor kinase that stimulates GacS–GacA activity and promotes biofilm features[42], yet this strain still established chronic localized infection in the mouse chronic wound model (Fig. 3b). Like many other pathogenic bacteria, *P. aeruginosa* colonizes a wide range of ecological niches including host-free ecosystems, and as a result, certain regulatory features controlling the biofilm–planktonic transition may have been selected under conditions that are unrelated to human pathogenesis. Our study thus provides a framework for discovering important pathogenic traits by leveraging bacterial gene expression information from human infections.

Methods

Media and growth conditions

P. aeruginosa cells were routinely cultured in brain heart infusion (BHI) broth at 37 °C with shaking (250 r.p.m.) unless otherwise specified. *Pseudomonas* isolation agar (PIA) was used for selecting *P. aeruginosa* against other microbes. *Escherichia coli* cells were cultured in lysogeny broth. To prepare plates, 1.5% (w/v) agar was used. Antibiotics were added to the medium as necessary: 60 µg ml^{-1} gentamicin, 75 µg ml^{-1} Tet or 300 µg ml^{-1} carbenicillin for *P. aeruginosa*; 10 µg ml^{-1} gentamicin, 10 µg ml^{-1} Tet or 100 µg ml^{-1} carbenicillin for *E. coli*. For *P. aeruginosa* static or anaerobic cultivation, 50 mM KNO_3 was added to BHI broth. The anaerobic condition was maintained in a vinyl anaerobic chamber (Coy Lab Products) with the following atmosphere: 85% N_2, 10% CO_2 and 5% H_2.

Plasmids and bacterial strains

Strains, plasmids and primers used in this study are listed in Supplementary Table 3. To create transcriptional *lacZ* fusions (pPC100–102), desired promoter regions were PCR amplified (Q5, New England Biolabs) using PA14 genomic DNA as the template, and the resulting amplicons containing overlapping regions of about 25 base pairs (bp) were ligated into the mini-

CTX-*lacZ* vector (linearized with EcoRI and KpnI) in Gibson Assembly Master Mix (New England Biolabs). To create *sicX* complementation plasmids (pPC103–106), the *sicX* promoter and *sicX* alleles (containing mutations as described in Supplementary Table 3) were ligated into mini-CTX1 vector (linearized with EcoRI and KpnI) through Gibson assembly. All *sicX* point mutations (pPC107–124) were constructed with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs), using mini-CTX1-*sicX*$^{\text{Truncate}}$ (pPC106) as the template. To create translational *lacZ* fusions (pPC125–126), the promoter region of *ubiU* was PCR amplified and then ligated into pSW205 between the EcoRI and BamHI sites. All *ubiUVT* 5′ UTR point mutations (pPC127–145) were constructed with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs), using pSW205-P_{ubi}-*lacZ* (pPC125) as the template. To construct plasmids used for markerless gene deletion (pPC146–147, 149–150) and mutagenesis (pPC148), upstream and downstream regions (about 1,000 bp) flanking the target gene or region were PCR amplified (containing mutations in the overlapping region if needed) and ligated into pEXG2 (linearized with SacI and KpnI) through Gibson assembly. All plasmids were verified by PCR, restriction enzyme digestion and, if necessary, DNA sequencing.

Verified plasmids were then transformed into *E. coli* SM10 λ*pir*, followed by conjugation into *P. aeruginosa* PA14, PAO1, B1 and their derivatives. The pSW205-derived plasmids were directly electroporated into *P. aeruginosa*. Appropriate antibiotics were used for selection as indicated in Supplementary Table 3. To create markerless in-frame deletion or mutagenesis, conjugants with the deletion cassette integrated at the target loci were counter-selected on low-salt lysogeny broth agar plates supplemented with 5% sucrose. The successful deletion was screened by PCR using flanking primers and verified by DNA sequencing if needed.

Homologue identification and sequence analysis

Blastn searches were carried out on pseudomonas.com and the National Center for Biotechnology Information website to identify orthologues of *sicX* and *ubiUVT* across the *Pseudomonas* genus and other organisms. The *sicX* sequence (for Fig. 1c and Extended Data Fig. 1), *sicX* with its neighbouring genes (Fig. 1d) and *ubiUVT* with its upstream intergenic

region (for Extended Data Fig. 4) from the *P. aeruginosa* PA14 strain were used as queries. Orthologues identified from complete bacterial genomes with *E* values of $<1 \times 10^{-4}$, percentage identity >90% and query coverage values of >90% were retained for further analysis. Sequences of *sicX* and *ubiUVT* orthologues were aligned using MUSCLE[43]. To analyse the sequence conservation of *sicX* and *ubiUVT* orthologues, we first aligned *sicX* and *ubiUVT* orthologues using MUSCLE. Next, alignment ends were trimmed and gaps were removed, and percentages of orthologue sequences that are identical to *sicX* and *ubiUVT* queries were calculated at each nucleotide position in R. To create Extended Data Fig. 4b,c, the alignments were visualized in Jalview[44] using the default nucleotide colour scheme.

β-galactosidase assay

P. aeruginosa cells were grown overnight in BHI broth supplemented with 50 mM KNO_3. The overnight culture was diluted to an optical density at 600 nm (OD_{600nm}) of about 0.05 in fresh BHI broth (with 50 mM KNO_3) and then grown to the log phase (OD_{600nm} around 0.5). Log-phase cells were diluted to OD_{600nm} 0.2 and transferred to the anaerobic chamber for 3-h shaking (at 100 r.p.m.) cultivation at 37 °C. A 100 µl volume of the anaerobic culture was used to conduct the β-galactosidase assay (as described in ref. [45]). Aerobic log-phase cultures were used as comparisons. For static cultivation, log-phase cultures were diluted to OD_{600nm} 0.05 and incubated in a 96-well plate (500 µl in each 1-ml-deep well) statically at 37 °C for 3 h. The static cultures were then used for the β-galactosidase assay.

For making Extended Data Fig. 3a,b, strains harbouring *ubiUVT-lacZ* translational fusions were spotted (3 µl of OD_{600nm} 0.1 cultures) onto 0.2-µm Nuclepore track-etched polycarbonate membranes (Whatman) that were placed (rough side facing upwards) on BHI agar plates containing 200 µg ml^{-1} X-gal, 50 mM KNO_3 and 100 µg ml^{-1} carbenicillin. The plates were then incubated in the anaerobic chamber (protected from light) at 37 °C for 20 h before imaging. β-galactosidase activities were estimated as described previously[46]. Briefly, images of colony biofilms were converted to 8 bit in ImageJ. Thresholding was carried out to define colony objects for

measurements. Mean pixel intensity was measured for each colony and subtracted with the mean pixel intensity of a colony of WT PA14.

Northern blot

P. aeruginosa cells were grown aerobically and anaerobically for total RNA extraction. To prepare aerobic cultures, log-phase cells were diluted to OD_{600nm} 0.01 in BHI (with 50 mM KNO_3) and incubated at 37 °C with shaking for 3 h. To prepare anaerobic cultures, log-phase cells were diluted to OD_{600nm} 0.1 in BHI (with 50 mM KNO_3) and incubated in the anaerobic chamber for 4 h with shaking at 37 °C. For total RNA extraction, *P. aeruginosa* cultures were first stored in RNAlater at 4 °C overnight. Next, cells were pelleted and resuspended in RNase-free TE buffer containing 1 mg ml^{-1} lysozyme. Samples were incubated at 37 °C for 30 min to enzymatically lyse cells. A 1 ml volume of RNA-Bee was then added, and cells were further mechanically lysed by bead beating three times for 30 s. A 200 µl volume of chloroform was added to the samples, and the aqueous and organic phases were separated by centrifugation at 13,000*g* for 30 min at 4 °C. The aqueous phase that contained RNA was mixed with 0.5 ml isopropanol and 20 µg linear acrylamide. After overnight incubation at −80 °C, RNA was pelleted and washed with 75% ethanol. Air-dried pellets were resuspended in 100 µl of RNase-free water. The RNA concentration for each sample was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific). Approximately 1 µg of total RNA was separated on a 15% polyacrylamide gel electrophoresis–urea gel and transferred to Hybond-N+ membranes (GE Healthcare). The membranes were then crosslinked with an ultraviolet crosslinker (UVP) and probed in hybridization solution (Takara) with the indicated DNA oligonucleotides (listed in Supplementary Table 3) that were radiolabelled with [γ-^{32}P]ATP (Perkin Elmer) by T4 polynucleotide kinase (New England Biolabs). Probed membranes were exposed to a phosphor screen and visualized using a Typhoon Biomolecular Imager (GE Healthcare). When necessary, membranes were stripped by incubating them with boiled 0.1% SDS with agitation for 15 min three times. The sizes of RNA were estimated with the RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific) as well as the migration of xylene cyanol (about 28 nt) and bromophenol blue (about 8 nt).

Colony biofilm assay

Bacterial cultures in BHI (with 50 mM KNO_3) were grown to the log phase (OD_{600nm} around 0.5) and then diluted to OD_{600nm} 0.02 with fresh medium. Next, 10 µl diluted cultures were spotted onto 0.2-µm Nuclepore track-etched polycarbonate membranes (Whatman) that were placed (smooth side facing upwards) on BHI agar plates (with 50 mM KNO_3). The plates were then incubated in the anaerobic chamber at 37 °C for 16 h. Membranes containing *P. aeruginosa* biofilms were transferred to the BeadBug prefilled tubes that contained 1 ml phosphate-buffered saline (PBS). Biofilms were mechanically disrupted by bead beating for 45 s. CFUs were determined by spreading the diluted cell suspensions on PIA plates.

Quantitative proteomics and data analysis

P. aeruginosa PA14 and Δ*sicX* cells were collected under both static and anaerobic growth conditions as described above (in the section entitled β-galactosidase assay).

For static growth condition, three biological replicates for each strain were collected. Cell pellets were resuspended in lysis buffer (100 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, 1 mM Tris(2-carboxyethyl)phosphine) (TCEP), pH 7.5) with cOmplete Mini protease inhibitor cocktail (Roche). Following three rounds of sonication (3 × 10 s) on ice, supernatants were clarified by sedimentation (21,130*g*, 15 min, 4 °C). Aliquots (100 µg) of each sample were reduced with TCEP, alkylated with iodoacetamide and labelled with tandem mass tags (TMTs). TMT labelling was carried out according to the protocol specified by the manufacturer (Thermo Fisher).

Liquid chromatography–tandem mass spectrometry (LC–MS/MS) experiments were carried out using a Dionex Ultimate 3000 rapid separation LC (RSLC) nano-ultraperformance LC system (Thermo Fisher Scientific) and a Lumos Orbitrap mass spectrometer (Thermo Fisher Scientific). Separation of peptides was carried out by C18 reversed-phase chromatography at a flow rate of 300 nl min^{-1} using a Thermo Scientific reversed-phase nano EASY-spray column (Thermo Scientific PepMap C18; 2-mm particle size, 100-Å pore size, 75-mm i.d. by 50-cm length). Solvent

A was water/0.1% formic acid, and solvent B was 80% (v/v) acetonitrile/20% water/0.1% formic acid. The linear gradient used was 2% to 40% solvent B over 93 min (the total LC run time was 120 min, including a high-organic-wash step and column re-equilibration).

The eluted peptides were sprayed into the mass spectrometer by means of an EASY-Spray source (Thermo Fisher Scientific). All *m*/*z* values representing eluting peptide ions were measured in an Orbitrap mass analyser (set at a resolution of 120,000) and were scanned at between *m*/*z* 380 and 1,500 Da. Data-dependent MS/MS scans (top speed) were used to automatically isolate and fragment precursor ions by collision-induced dissociation (normalized collision energy value, 35%) analysed in the linear ion trap. Singly charged ions and ions with unassigned charge states were excluded from selection for MS/MS, and a dynamic exclusion window of 70 s was used. The top 10 most abundant fragment ions from each MS/MS event were then selected for a further stage of fragmentation by synchronous precursor selection and MS/MS/MS[47] in the high-energy-collision cell using high-energy collisional dissociation (normalized collision energy value of 65%). The *m*/*z* values and relative abundances of each reporter ion and of all fragments (mass range, 100 to 500 Da) in each MS3 step were measured in the Orbitrap analyser, which was set at a resolution of 60,000. This was carried out in cycles of 10 MS3 events, after which the Lumos instrument reverted to scanning the *m*/*z* ratios of the intact peptide ions and the cycle continued.

Proteome Discoverer v2.1 (Thermo Fisher Scientific) and Mascot (Matrix Science) v2.6 were used to process raw data files. Data were aligned to the sequences of *P. aeruginosa* UCBPP-PA14 (with common repository of adventitious proteins (cRAP) v1.0). The R package MSnbase v2.13 (ref. [48]) was used for processing proteomics data. Protein differential abundances were evaluated using the Limma package v3.44.3 (ref. [49]). Differences in protein abundances were statistically determined using Student's *t*-test with variances moderated by the use of Limma's empirical Bayes method. *P* values were adjusted for multiple testing by the Benjamini–Hochberg method[50]. Proteins were considered to have increased or decreased in abundance only when their $\log_2$[fold change] value was greater than 1 or less than −1, respectively, and when their *P* value was <0.01.

For anaerobic growth condition, two biological replicates for PA14 and three biological replicates for Δ*sicX* were collected as described above (in the section entitled β-galactosidase assay). Briefly, the proteins in each sample were reduced, alkylated and digested with trypsin according to the filter-aided sample preparation protocol[51]. The peptides were labelled with TMTs, separated by high-pH reversed-phase chromatography as described previously[52]. They were pooled into 12 fractions as described previously[53]. Each fraction was analysed by nano-LC–MS/MS, and peptides were identified as previously described[54] with the following modifications. Reversed-phase chromatography was carried out using an in-house packed column (40 cm long × 75 µm inner diameter × 360 µm outer diameter, Dr. Maisch GmbH ReproSil-Pur 120 C18-AQ 1.9-µm beads) and a 120-min gradient. The Raw files were searched using the Mascot algorithm (v2.5.1) against a protein database constructed by combining the sequences of *P. aeruginosa* UCBPP-PA14 and a contaminant database (cRAP, downloaded 21 November 2016 from http://www.thegpm.org) through Proteome Discoverer v2.1. Only peptide spectral matches with expectation value of less than 0.01 ('high confidence') were used for our analyses as described above.

Lipid extractions and ubiquinone (UQ9) quantification

To prepare cells for lipid extractions, we first diluted log-phase cultures (OD_{600nm} 0.5; grown aerobically) into the anaerobic BHI broth (supplemented with 50 mM KNO_3) to a calculated OD_{600nm} 0.05, and after 6-h anaerobic incubation with shaking, cells were collected, washed once with PBS and flash-frozen in liquid nitrogen. Cell numbers were estimated with OD_{600nm}.

Frozen samples were thawed on ice. A 100 µl volume of ice-cold isopropanol with 10 nM coenzyme Q10-d_9 (deuterated UQ10 as internal standard; Sigma-Aldrich) was added to 10^9 cells. With the addition of glass beads, samples were vortexed briefly and homogenized by TissueLyzer for 5 min twice, followed by centrifugation at 21,100*g* for 5 min. After centrifugation, supernatant was diluted tenfold and transferred to 4 °C. Ultraperformance LC–MS was carried out using an UltiMate 3000 fitted

with an Accucore C30 column (2.1 × 150 mm, 2.6 µm particle size; Thermo Fisher) and coupled to an Orbitrap ID-X mass spectrometry system.

The chromatographic method for sample analysis involved elution with acetonitrile/water (60:40, v/v) with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and isopropanol/acetonitrile (90:10, v/v) with 10 mM ammonium formate and 0.1% formic acid (mobile phase B) at 0.4 ml min^{-1} flow rate using the following gradient programme: 0 min 20% B; 1 min 60% B; 5 min 70% B; 5.5 min 85% B; 8 min 90% B; 8.2 min 100% B hold to 10.5 min, then 10.7 min 20% B hold to 12 min. The column temperature was set to 50 °C, and the injection volume was 5 µl.

The targeted molecules UQ9 and deuterated UQ10 were fragmented by high-energy collisional dissociation at 20 collision energy in positive mode. MS/MS transitions for UQ9 and deuterated UQ10 are 812.66/197.08 and 889.77/206.18. Standard curves for UQ9 and deuterated UQ10 were generated. The amounts of UQ9 and deuterated UQ10 in the samples were calculated on the basis of the standard curve.

Planktonic growth and competition assays

For generating growth curves, bacterial cells in BHI (with 50 mM KNO_3) were first grown aerobically to the log phase (OD_{600nm} around 0.5) and then diluted to OD_{600nm} 0.01 with either aerobic or anaerobic BHI (with 50 mM KNO_3) broth. Aerobic and anaerobic growth of *P. aeruginosa* at 37 °C (with shaking) were monitored by measuring the OD_{600nm}.

For long-term competition experiments, PA14, Δ*sicX*, Tet^R (resistance provided by mini-CTX1) PA14 and Tet^R Δ*sicX* were grown to the log phase (OD_{600nm} around 0.5) in BHI (with 50 mM KNO_3) before diluting to OD_{600nm} 0.4 with fresh medium. PA14 and Tet^R Δ*sicX* were mixed at 1:1 ratio, and 100 µl of cell mixtures was transferred to 4 ml anaerobic BHI (with 50 mM KNO_3), followed with shaking incubation at 37 °C under anaerobic conditions. After 24 h, 20 µl of the culture was transferred to another tube containing 4 ml anaerobic BHI (with 50 mM KNO_3). The culture was then consecutively passaged daily in a similar fashion. The

competition between Tet^R PA14 and Δ*sicX* was conducted similarly to that described above. CFU values of Tet^R and Tet^S cells were estimated by plating cultures (on each day) on PIA as well as PIA containing 75 μg ml^{-1} Tet.

In silico analyses of RNA folding and RNA–RNA interaction

RNA secondary structures were predicted using mfold[26] with default parameters. IntaRNA 2.0 was used to predict base-pairing interactions between SicX sRNA and *ubiUVT* mRNA[25] with default settings, and the minimum number of base pairs in the seed region was set as 7.

Mouse chronic wound model

Mouse chronic wound infections were carried out with female Swiss Webster mice (8–10 weeks old) essentially as described previously[55] with a few modifications. Briefly, dorsal full-thickness skin excision was carried out in which a small circular wound (with a diameter of 1.5 cm) was surgically administered. After excision, the wound was immediately covered with a semipermeable polyurethane dressing. A total of 10^4 CFUs of *P. aeruginosa* cells were injected onto the wound bed underneath the dressing to establish infection. As bandages are required to reduce the contractile healing and minimize wound contamination with other bacteria, they were closely monitored and replaced when necessary throughout the infection course.

For assessing the bacterial burdens of WT and Δ*sicX* infection, wound tissues and spleens were collected at 10 days post-infection or at an early time point if the animal was found moribund (moribund mice that were not identified immediately were excluded from the CFU analysis). Tissues were homogenized in PBS for 45 s in BeadBug tubes with 2.8-mm steel beads (Sigma-Aldrich) using a Mini-Beadbeater-16 (BioSpec Products). To enumerate CFUs, homogenized samples were serially diluted and plated on PIA plates. For studying the dissemination outcomes of WT, Δ*sicX*, Δ*ubiUVT* and Δ*sicX*-*ubi*^{C} infection (Fig. 3e), at 10 days post-infection, distal organs including spleens, livers and gallbladders were collected for assessing the presence or absence of *P. aeruginosa* cells. For studying the

macroscale organization of *P. aeruginosa* infection, the wound was excised at 4 days post-infection, and a biopsy punch of 10 mm diameter (Acu-Punch, Thermo Fisher Scientific) was used to separate the edge (about 0.98 cm^2) and core (about 0.77 cm^2) regions. The CFUs were then normalized by the surface area for comparison purposes.

In vivo biofilm dispersal

In vivo biofilm dispersal was carried out with female Swiss Webster mice (8–10 weeks old) essentially as described previously[31] with a few modifications. Briefly, the mouse surgical excision wound was administered as demonstrated above. A total of 10^4 CFUs of *P. aeruginosa* cells (PAO1) were injected onto the wound bed underneath the dressing to establish infection. At 48 h post-infection (after the formation of mature biofilms in wounds), the established wound infections were treated through topical application of GHs or *cis*-DA to induce systemic dissemination. Specifically, GHs are composed of bacterial α-amylase (from *Bacillus subtilis*; MP Biomedicals) and fungal cellulase (from *Aspergillus niger*; MP Biomedicals). A 10% (w/v) α-amylase and cellulase (in a 1:1 combination) solution was prepared by dissolving powdered enzymes in PBS. *cis*-DA (Cayman Chemical Company) was diluted in PBS to a final concentration of 500 nM. GH and *cis*-DA solutions were prepared immediately before use. Wound beds were irrigated with a GH or *cis*-DA solution in three separate topical infusions with 30 min of dwell time for each. Topical infusions containing dispersed cells ($n = 2$ animals for each treatment) were aspirated and immediately placed in RNAlater (Thermo Fisher Scientific). In separate animals, *P. aeruginosa*-infected wound tissues ($n = 3$ animals) were carefully excised from the surrounding uninfected tissue and the underneath muscle layer at 48 h post-infection (without GH or *cis*-DA treatment) as a comparison. Infected wound tissues were immediately placed in RNAlater (Thermo Fisher Scientific).

Mouse pneumonia model

Mouse pneumonia infection was carried out with female BALB/c mice (6–8 weeks old) essentially as described previously[56] with a few modifications. Briefly, mice were anaesthetized by intraperitoneal injection of 0.2 ml of a

cocktail of ketamine (25 mg ml^{-1}) and xylazine (12 mg ml^{-1}). For infection with PA14-derived strains, mice were intranasally instilled with approximately 10^7 CFUs (a sublethal dose) *P. aeruginosa* cells (in 25 μl PBS). Mice were euthanized at 48 h post-infection. Whole lungs and spleens were collected aseptically, weighed and homogenized in 1 ml of PBS. Tissue homogenates were serially diluted and plated on PIA for CFU enumeration. For infection with PAO1-derived strains, 5×10^7 CFUs (a lethal dose) of *P. aeruginosa* cells were used as inocula, and tissues were collected at 24 h post-infection.

Mouse subcutaneous infection model

Mouse subcutaneous skin infection (also known as the mouse abscess model) was carried out with female Swiss Webster mice (8 weeks old) essentially as described previously[55] with a few modifications. Briefly, the mouse inner thigh was shaved and any remaining hair was removed with a depilatory agent (Nair). A 100 μl volume of approximately 1×10^7 CFUs of *P. aeruginosa* cells (PA14 or Δ*sicX*) were subcutaneously injected to the inner thigh. Mice were euthanized at 16 h post-infection. Spleens were collected aseptically and homogenized in 1 ml PBS. Tissue homogenates were serially diluted and plated on PIA for CFU enumeration.

RNA-seq library preparation

RNA extractions were carried out essentially as described previously[14]. For in vitro static cultivation of PA14, cells were first grown as described above (in the section entitled β-galactosidase assay) and immediately mixed with five volumes of RNAlater. After storage at 4 °C overnight, cells were pelleted, resuspended in nuclease-free TE buffer (Acros Organics) containing 1 mg ml^{-1} lysozyme, and transferred to bead-beating tubes containing a mixture of large and small beads (2-mm zirconia and 0.1-mm zirconia/silica, respectively). For in vivo biofilm dispersal experiments (as described above), infected tissues were removed from RNAlater, mixed with TE buffer, and transferred to bead-beating tubes. In vitro as well as tissue samples in TE were briefly disrupted by homogenization using a Mini-Beadbeater-16. Samples were then incubated at 37 °C for 30 min to enzymatically lyse cells. Next, 1 ml RNA-Bee was added to 300-μl sample

homogenate, and samples were further mechanically lysed by bead beating three times for 30 s. A 200 µl volume of chloroform per 1 ml RNA-Bee was added, and the tubes were shaken vigorously for 30 s and then incubated on ice for 5 min. The samples were then centrifuged at 13,000*g* for 30 min at 4 °C to separate the aqueous and organic phases. The aqueous phase was transferred to a new microcentrifuge tube to which 0.5 ml isopropanol (per 1 ml RNA-Bee) and 20 µg linear acrylamide were added. After overnight incubation at −80 °C, samples were thawed on ice and centrifuged at 13,000*g* for 30 min at 4 °C. Pellets were washed twice with 1 ml of 75% ethanol, air dried for 5 min and resuspended in 100 µl of RNase-free water. The RNA concentration for each sample was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific). rRNA was depleted using a MICROBExpress Kit (Thermo Fisher Scientific) for the in vitro samples and a QIAseq FastSelect Kit (QIAGEN) for mouse-derived samples, followed by ethanol precipitation. The depleted RNA was fragmented for 2 min with the NEBNext Magnesium RNA fragmentation module kit, and cDNA libraries were prepared using the NEBNext multiplex small RNA library prep kit (New England Biolabs) according to the manufacturer's instructions. Libraries were sequenced at the Molecular Evolution Core at the Georgia Institute of Technology on an Illumina NextSeq500 using 75-bp single-end runs.

Analysis of RNA-seq datasets

In addition to the above-mentioned RNA-seq libraries prepared in this study, 202 publicly available *P. aeruginosa* RNA-seq datasets were analysed, including 28 *P. aeruginosa* transcriptomes derived from human samples, 28 from infection models and 146 from a range of well-defined in vitro conditions (listed in Supplementary Table 1). The quality of the raw sequencing reads was first confirmed with FastQC v0.11.7 (ref. [57]). Next, RNA-seq reads were trimmed on the 3′ ends to remove the Illumina adaptor (AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C) using Cutadapt 3.0 (ref. [58]) with a minimum read length threshold of 22 bases. Trimmed reads were then mapped to the *P. aeruginosa* PA14 reference genome (available for download from pseudomonas.com) using Bowtie2 v2.4.2 with default parameters for end-to-end alignment[59]. As *PA1414* (*sicX*) was originally annotated as a protein-coding gene, we first assigned reads to

protein-coding genes with featureCounts Subread v2.0.1 (ref. [60]; results shown in Fig. 1a). For the remaining analyses (Figs. 1b, 2a and 4b–d and Extended Data Figs. 2b and 10), reads assigned to both protein-coding genes as well as 199 sRNAs[15] were tallied.

Differential expression was determined with DESeq2 (Figs. 1a and 4b–d; ref. [61]). To compare *sicX* expression levels in different samples (Fig. 1a and Extended Data Fig. 5), raw reads were normalized using the varianceStabilizingTransformation() function in the DESeq2 package. To estimate the relative transcript abundance of various features (protein-coding genes, sRNAs, *sicX* and/or *rsmYZ*), we calculated TPM. First, raw counts were normalized by feature (gene) length to generate reads per kilobase (RPK). Next, all RPK values were counted up and divided by 1,000,000 to determine the scaling factor. Finally, RPK of each gene was divided by the scaling factor. The resulting TPM values were used for making Figs. 1b and 2a and Extended Data Figs. 2b and 10. To examine the alignment pattern of *sicX* reads, samtools v1.13 was used to measure the read depth encompassing the *sicX* locus at each nucleotide position[62]. The read depth was then normalized using the estimateSizeFactors() function in DESeq2 before plotting.

Phylogenetic analysis

A maximum-likelihood phylogenetic tree was constructed in PhyML 3.0 (ref. [63]) using 365 16S rRNA sequences representing diverse *Pseudomonas* isolates. For better visualization of the tree, we included only approximately 50 16S sequences from the *P. aeruginosa* strains. The 16S sequences were aligned with MUSCLE and used as an input in PhyML 3.0. The best-fitting evolutionary models were predicted with the corrected Akaike information criterion. The tree was visualized and annotated in iTOL[64].

Functional enrichment analysis

Gene Ontology enrichment analysis was conducted using Galaxy v2.0.1. Differentially expressed genes in response to GH or *cis*-DA treatment were analysed using the Gene Ontology file (obtained from http://geneontology.org/docs/download-ontology) and the *P. aeruginosa*

PAO1 Gene Ontology term annotations (obtained from https://pseudomonas.com/goterms/list). A Benjamini–Hochberg test with *P*-value cutoff < 0.05 was carried out.

Statistics and reproducibility

All experiments were carried out independently at least three times (unless otherwise stated in the figure legends) with similar observations. The exact *n* values for animal studies as well as in vitro experiments are provided in the figures and legends. Statistical analyses were conducted using Prism GraphPad 9 and are specified in the figure legends. For northern blotting, each experiment was conducted at least twice independently (by using freshly extracted RNA samples) with similar observations, and representative images are shown.

Ethics statement

All mouse procedures were carried out in strict accordance with the recommendations in the *Guide for the Care and Use of Laboratory Animals* of the National Institutes of Health. All animals were treated humanely and housed in temperature- and humidity-controlled facilities (18–22 °C; 40–50%) with 12 h light–dark cycles. The protocol for the mouse chronic wound infection was approved by the Institutional Animal Care and Use Committee (IACUC) of Texas Tech University Health Sciences Center (protocol number 07044) and by the IACUC of Georgia Institute of Technology (protocol number A100127). The protocol for the mouse subcutaneous infection model was approved by the IACUC of Georgia Institute of Technology (protocol number A100127). For the mouse pneumonia infection model, all experimental procedures were conducted according to the guidelines of the Emory University IACUC, under approved protocol number DAR-201700441.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The proteomics datasets generated in this study have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org/cgi/GetDataset), with the accession numbers are PXD032985 and PXD032986. Raw reads of our RNA-seq studies have been deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject numbers PRJNA934328 and PRJNA904394. The 202 publicly available *P. aeruginosa* transcriptomes analysed in this study are summarized in Supplementary Table 1, and can be obtained from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the corresponding accession number. The *P. aeruginosa* PA14 reference genome (BioSample accession SAMN02603591) is available for download from https://pseudomonas.com. Source data are provided with this paper.

References

1. Rossi, E. et al. *Pseudomonas aeruginosa* adaptation and evolution in patients with cystic fibrosis. *Nat. Rev. Microbiol.* **19**, 331–342 (2021).

2. Valentini, M., Gonzalez, D., Mavridou, D. A. & Filloux, A. Lifestyle transitions and adaptive pathogenesis of *Pseudomonas aeruginosa*. *Curr. Opin. Microbiol.* **41**, 15–20 (2018).

3. Valentini, M. & Filloux, A. Biofilms and cyclic di-GMP (c-di-GMP) signaling: lessons from *Pseudomonas aeruginosa* and other bacteria. *J. Biol. Chem.* **291**, 12547–12555 (2016).

4. Ha, D.-G. & O'Toole, G. A. c-di-GMP and its effects on biofilm formation and dispersion: a *Pseudomonas aeruginosa* review. *Microbiol. Spectr.* **3**, 3.2.27 (2015).

5. Pestrak, M. J. & Wozniak, D. J. in *Microbial Cyclic Di-Nucleotide Signaling* (eds Chou, S.-H., Guiliani, N., Lee, V. T. & Römling, U.) 471–486 (Springer Cham, 2020).

6. Brencic, A. et al. The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Mol. Microbiol.* **73**, 434–445 (2009).

7. Goodman, A. L. et al. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**, 745–754 (2004).

8. Gebhardt, M. J., Kambara, T. K., Ramsey, K. M. & Dove, S. L. Widespread targeting of nascent transcripts by RsmA in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **117**, 10520–10529 (2020).

9. Janssen, K. H., Diaz, M. R., Gode, C. J., Wolfgang, M. C. & Yahr, T. L. RsmV, a small noncoding regulatory RNA in *Pseudomonas aeruginosa* that sequesters RsmA and RsmF from target mRNAs. *J. Bacteriol.* **200**, e00277-18 (2018).

10. Miller, C. L. et al. RsmW, *Pseudomonas aeruginosa* small non-coding RsmA-binding RNA upregulated in biofilm versus planktonic growth conditions. *BMC Microbiol.* **16**, 155 (2016).

11. Broder, U. N., Jaeger, T. & Jenal, U. LadS is a calcium-responsive kinase that induces acute-to-chronic virulence switch in *Pseudomonas aeruginosa*. *Nat. Microbiol.* **2**, 16184 (2016).

12. Wang, B. X. et al. Mucin glycans signal through the sensor kinase RetS to inhibit virulence-associated traits in *Pseudomonas aeruginosa*. *Curr. Biol.* **31**, 90–102 (2021).

13. Cornforth, D. M. et al. *Pseudomonas aeruginosa* transcriptome during human infection. *Proc. Natl Acad. Sci. USA* **115**, E5125–E5134 (2018).

14. Cornforth, D. M., Diggle, F. L., Melvin, J. A., Bomberger, J. M. & Whiteley, M. Quantitative framework for model evaluation in microbiology research using *Pseudomonas aeruginosa* and cystic fibrosis infection as a test case. *MBio* **11**, e03042-19 (2020).

15. Trouillon, J., Han, K., Attrée, I. & Lory, S. The core and accessory Hfq interactomes across *Pseudomonas aeruginosa* lineages. *Nat. Commun.* **13**, 1258 (2022).

16. Trunk, K. et al. Anaerobic adaptation in *Pseudomonas aeruginosa*: definition of the Anr and Dnr regulons. *Environ. Microbiol.* **12**, 1719–1733 (2010).

17. Rossi, E., Falcone, M., Molin, S. & Johansen, H. K. High-resolution in situ transcriptomics of *Pseudomonas aeruginosa* unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Nat. Commun.* **9**, 3459 (2018).

18. Schobert, M. & Tielen, P. Contribution of oxygen-limiting conditions to persistent infection of *Pseudomonas aeruginosa*. *Fut. Microbiol.* **5**, 603–621 (2010).

19. Huse, H. K. et al. Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations in vivo. *mBio* **1**, e00199-10 (2010).

20. Nelson, C. E. et al. Proteomic analysis of the *Pseudomonas aeruginosa* iron starvation response reveals PrrF small regulatory RNA-dependent iron regulation of twitching motility, amino acid metabolism, and zinc homeostasis proteins. *J. Bacteriol.* **201**, e00754-18 (2019).

21. Han, K. & Lory, S. Toward a comprehensive analysis of posttranscriptional regulatory networks: a new tool for the identification of small RNA regulators of specific mRNAs. *mBio* **12**, e03608-20 (2021).

22. Pusic, P., Sonnleitner, E. & Bläsi, U. Specific and global RNA regulators in *Pseudomonas aeruginosa*. *Int. J. Mol. Sci.* **22**, 8632 (2021).

23. Vo, C.-D.-T. et al. The O_2-independent pathway of ubiquinone biosynthesis is essential for denitrification in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **295**, 9021–9032 (2020).

24. Rompf, A. et al. Regulation of *Pseudomonas aeruginosa hemF* and *hemN* by the dual action of the redox response regulators Anr and Dnr. *Mol. Microbiol.* **29**, 985–997 (1998).

25. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).

26. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

27. Updegrove, T. B., Zhang, A. & Storz, G. Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).

28. Sagawa, S., Shin, J.-E., Hussein, R. & Lim, H. N. Paradoxical suppression of small RNA activity at high Hfq concentrations due to random-order binding. *Nucleic Acids Res.* **43**, 8502–8515 (2015).

29. Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S. A. Positive regulation by small RNAs and the role of Hfq. *Proc. Natl Acad. Sci. USA* **107**, 9602–9607 (2010).

30. Castillo-Keller, M., Vuong, P. & Misra, R. Novel mechanism of *Escherichia coli* porin regulation. *J. Bacteriol.* **188**, 576–586 (2006).

31. Redman, W. K., Welch, G. S. & Rumbaugh, K. P. Assessing biofilm dispersal in murine wounds. *J. Vis. Exp.* **174**, e62136 (2021).

32. Sauer, K. et al. Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *J. Bacteriol.* **186**, 7312–7326 (2004).

33. Naumoff, D. Hierarchical classification of glycoside hydrolases. *Biochemistry* **76**, 622–635 (2011).

34. Fleming, D., Chahin, L. & Rumbaugh, K. Glycoside hydrolases degrade polymicrobial bacterial biofilms in wounds. *Antimicrob. Agents Chemother.* **61**, e01998-16 (2017).

35. Fleming, D. & Rumbaugh, K. The consequences of biofilm dispersal on the host. *Sci. Rep.* **8**, 10738 (2018).

36. Davies, D. G. & Marques, C. N. A fatty acid messenger is responsible for inducing dispersion in microbial biofilms. *J. Bacteriol.* **191**, 1393–1403 (2009).

37. Marques, C. N., Davies, D. G. & Sauer, K. Control of biofilms with the fatty acid signaling molecule *cis*-2-decenoic acid. *Pharmaceuticals* **8**, 816–835 (2015).

38. Rahmani-Badi, A., Sepehr, S., Fallahi, H. & Heidari-Keshel, S. Dissection of the *cis*-2-decenoic acid signaling network in *Pseudomonas aeruginosa* using microarray technique. *Front. Microbiol.* **6**, 383 (2015).

39. Westermann, A. J. et al. Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **529**, 496–501 (2016).

40. Kinoshita-Daitoku, R. et al. A bacterial small RNA regulates the adaptation of *Helicobacter pylori* to the host environment. *Nat. Commun.* **12**, 2085 (2021).

41. Mulcahy, H. et al. *Pseudomonas aeruginosa* RsmA plays an important role during murine infection by influencing colonization, virulence, persistence, and pulmonary inflammation. *Infect. Immun.* **76**, 632–638 (2008).

42. Mikkelsen, H., McMullan, R. & Filloux, A. The *Pseudomonas aeruginosa* reference strain PA14 displays increased virulence due to a mutation in *ladS*. *PLoS ONE* **6**, e29113 (2011).

43. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

44. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

45. Pearson, J. P. et al. Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. *Proc. Natl Acad. Sci. USA* **91**, 197–201 (1994).

46. Mould, D. L., Botelho, N. J. & Hogan, D. A. Intraspecies signaling between common variants of *Pseudomonas aeruginosa* increases production of quorum-sensing-controlled virulence factors. *MBio* **11**, e01865-20 (2020).

47. McAlister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).

48. Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288–289 (2012).

49. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer, 2005).

50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

51. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

52. Mertins, P. et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat. Protoc.* **13**, 1632–1661 (2018).

53. Yang, F., Shen, Y., Camp, D. G. 2nd & Smith, R. D. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Proteomics* **9**, 129–134 (2012).

54. Rinker, T. E. et al. Microparticle-mediated sequestration of cell-secreted proteins to modulate chondrocytic differentiation. *Acta Biomater.* https://doi.org/10.1016/j.actbio.2017.12.038 (2017).

55. Michie, K. L. et al. Role of *Pseudomonas aeruginosa* glutathione biosynthesis in lung and soft tissue infection. *Infect. Immun.* **88**, e00116-20 (2020).

56. Moustafa, D. A. et al. Peptide-conjugated phosphorodiamidate morpholino oligomers retain activity against multidrug-resistant *Pseudomonas aeruginosa* in vitro and in vivo. *MBio* **12**, e02411-20 (2021).

57. Andrews, S. FastQC v0.11.7 (Babraham Bioinformatics, Babraham Institute, 2010).

58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

60. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

62. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

63. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

64. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259

(2019).

Acknowledgements

We thank D. Cornforth and other members of the laboratory of M.W. for helpful discussion of the manuscript and assistance with the animal experiments. We thank G. Welch and H. Hui for help with the animal experiments. We thank S. Brown, S. Diggle and C. Baskett for critical reading of the manuscript. The proteomic and metabolomic studies were supported by the Georgia Institute of Technology Systems Mass Spectrometry Core Facility as well as the Cambridge Centre for Proteomics. This study was supported by National Institutes of Health grants R21AI154220 (to M.W.), R21AI137462 (to K.P.R.) and R21AI147178 (to C.S.S.), Cystic Fibrosis Foundation grants WHITEL20A0 and WHITEL22G0 (to M.W.) and Cystic Fibrosis Trust Foundation SRC017 (to M.W.). P.C. and S.K.D. are supported by Cystic Fibrosis Postdoctoral Fellowships CAO20F0 and DOLAN20F0, respectively.

Author information

Author notes

1. Derek Fleming

   Present address: Division of Clinical Microbiology, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

2. Whitni K. Redman

   Present address: Department of Biological Sciences, Binghamton University, Binghamton, NY, USA

3. These authors contributed equally: Stephen K. Dolan, Kayla H. Szymanik

Authors and Affiliations

1. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Pengbo Cao, Stephen K. Dolan, Anayancy Ramos, Frances L. Diggle & Marvin Whiteley

2. Emory-Children's Cystic Fibrosis Center, Atlanta, GA, USA

Pengbo Cao, Dina A. Moustafa, Stephen K. Dolan, Anayancy Ramos, Frances L. Diggle, Joanna B. Goldberg & Marvin Whiteley

3. Center for Microbial Dynamics and Infection, Georgia Institute of Technology, Atlanta, GA, USA

Pengbo Cao, Stephen K. Dolan, Anayancy Ramos, Frances L. Diggle & Marvin Whiteley

4. Department of Surgery, Texas Tech University Health Sciences Center, Lubbock, TX, USA

Derek Fleming, Whitni K. Redman & Kendra P. Rumbaugh

5. Department of Immunology and Molecular Microbiology, Texas Tech University Health Sciences Center, Lubbock, TX, USA

Derek Fleming, Whitni K. Redman & Kendra P. Rumbaugh

6. Burn Center of Research Excellence, Texas Tech University Health Sciences Center, Lubbock, TX, USA

Derek Fleming, Whitni K. Redman & Kendra P. Rumbaugh

7. Department of Pediatrics, Division of Pulmonary, Asthma, Cystic Fibrosis, and Sleep, Emory University School of Medicine, Atlanta, GA, USA

Dina A. Moustafa & Joanna B. Goldberg

8. Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA

Kayla H. Szymanik & Christopher S. Sullivan

Contributions

P.C. and M.W. conceived and designed the study, analysed data and wrote the manuscript; D.F., D.A.M., P.C., W.K.R. and F.L.D. contributed to various aspects of animal experiments; S.K.D. prepared samples and carried out data analyses for proteomics; K.H.S. carried out northern blotting and analyses; P.C., M.W. and A.R. constructed bacterial strains and carried out reporter assays; P.C. carried out all other experiments and data analysis; C.S.S., J.B.G. and K.P.R. provided technical guidance and analysed data.

Corresponding author

Correspondence to Marvin Whiteley.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 *PA1414* orthologs are restricted in *P. aeruginosa*.

The maximum-likelihood tree is built with 16s rDNA sequences from representative strains across the *Pseudomonas* genus. Clusters of abundant species are indicated with colored boxes. Strains that were reported to have been isolated from human or hospital environments are highlighted.

Extended Data Fig. 2 The expression of *PA1414* under low oxygen conditions.

a. Schematic of mutagenesis of the Anr binding motif. **b**. Figure 2a with the addition of five in vitro static growth transcriptomes (blue circles; the RNA-seq libraries were prepared following the same method used in our study of *P. aeruginosa* transcriptomes in humans). **c**. Planktonic growth (OD_{600}) of WT and Δ*PA1414* in the presence or absence of oxygen. Mean values from 3 independent experiments are shown. **d**. RNA-seq reads aligned to the *PA1414* locus (*sicX*) during in vitro aerobic and anaerobic growth. **e**. Colony biofilm growth of Δ*sicX* harboring different *sicX* mutations (as illustrated in Fig. 2g) under aerobic conditions. Error bars represent standard deviation from the mean. $n = 4$ independent experiments. Two-tailed Mann-Whitney test comparing WT and each other strain was performed (n.s., not significant). **f**. Northern blot analyses on different *sicX* alleles as depicted in Fig. 2g. The sRNA abundance was estimated by first normalizing against the 5S rRNA (loading control) and next comparing to the WT level. Anaerobic growth fitness was indicated with + (no defect) and − (defect). For gel source data, see Supplementary Fig. 1.

Extended Data Fig. 3 A proposed base-pairing interaction between SicX and the 5’ UTR of *ubiUVT*.

a. Predicted base-pairing between SicX and *ubiUVT* mRNA. Nucleotides highlighted in bold were subject to point mutations (A → U, U → A, G → C, C → G), which were assessed via the *ubiUVT-lacZ* translational reporter.

Colony biofilms grown under anaerobic conditions were assessed for β-galactosidase activities as described[46]. Heatmap displays the mean values. Nucleotides subject to compensatory base changes are highlighted with magenta borders. Predicted base-pairing regions are highlighted with dashed borders. **b**. β-galactosidase activities (individual values) of the experiment shown in (A) quantified as described[46]. Potential compensatory mutations are highlighted in blue. Green dashed lines indicate the mean β-galactosidase activities in WT and Δ*sicX*. A. U., arbitrary unit. $n = 5$ independent experiments. **c**. β-galactosidase assays evaluating whether 5' UTR mutations can alter *ubiUVT* translation independently of SicX during static growth. WT and Δ*sicX* harboring the *ubiUVT-lacZ* reporter (grown as anaerobic colony biofilms) were also assessed using this assay. $n \geq 3$ independent experiments. **d**. β-galactosidase assays indicating that the presence or absence of Hfq does not affect the induction of *ubiUVT* translation by SicX. $n \geq 5$ independent experiments. Error bars represent standard deviation from the mean. Two-tailed Mann-Whitney test was performed for **b**–**c** (n.s., not significant).

Extended Data Fig. 4 DNA sequence conservation of the *ubiUVT* operon.

a. The percent identity of 258 *ubiUVT* orthologs (compared to the *ubiUVT* allele from PA14) at the single-nucleotide scale. The *sicX* binding site at the 5' UTR is highlighted. **b**. DNA sequence alignment of the upstream intergenic region of 258 *ubiUVT* orthologs. **c**. DNA sequence alignment of 258 *sicX* orthologs (only the sRNA region is shown). The seed region in SicX and the target region in the 5' UTR of *ubiUVT* are indicated.

Extended Data Fig. 5 SicX expression in mouse chronic wounds.

a. RNA-seq reads aligned to the *PA1414* locus (*sicX*) in mouse chronic wound infection. **b**. *sicX* RNA level in the mouse chronic wound is comparable to that in humans. Lines indicate means. Two-tailed Mann-Whitney test was used to compare the bacterial burdens (n.s., not significant). **c**. Bacterial burdens in the wound (dissemination information shown in Fig. 3e). Lines indicate means. $n = 9$ animals (for each group) over three independent experiments. Two-tailed Mann-Whitney test was used to

compare the bacterial burdens of WT and each other mutant (n.s., not significant).

Source Data

Extended Data Fig. 6 WT and Δ*sicX* showed similar spleen colonization in a systemic infection model.

P. aeruginosa cells (~10^7 CFU) were subcutaneously injected into the mouse inner thigh to induce systemic infection. Bacterial burdens in spleens were assessed at 16 h post-infection. WT ($n = 7$) and Δ*sicX* ($n = 8$). Lines indicate means. Two-tailed Mann-Whitney test was used.

Source Data

Extended Data Fig. 7 Decoupling the translation of *ubiUVT* from SicX.

a. Schematic of mutagenesis in the 5’ UTR of *ubiUVT*: SicX binding site, ribosome binding site, and start codon are colored differently; the stem-loop structure is highlighted; mutagenesis (*ubi*^{C}) is indicated with an arrow. **b**. Restoring *ubiUVT* translation in Δ*sicX*. β-galactosidase assay was performed to assess the translational activity of *ubiUVT* with and without 5’ UTR mutation as described in A ($n = 6$). **c–d**. Growth yield of colony biofilms in the presence or absence of oxygen ($n = 4$). Error bars represent standard deviation from the mean. Significant differences (compared to the WT) are indicated with asterisks (*, $P < 0.05$; **, $P < 0.01$; Two-tailed Mann-Whitney test).

Extended Data Fig. 8 SicX (in PAO1 strain background) governs chronic-to-acute transition.

a. Schematic of the mouse pneumonia infection experiment. **b**. *P. aeruginosa* burden in the lung. Lines indicate means. **c**. *P. aeruginosa* burden in the spleen. Lines indicate means. Two-tailed Mann-Whitney test was used to compare bacterial burdens. Summary of two biological replicates for **b** and **c**.

Source Data

Extended Data Fig. 9 Functional enrichment of differentially-expressed genes in *P. aeruginosa* after GH or *cis*-DA treatment.

Gene Ontology (GO) terms are labeled on the left. One-tailed Fisher's exact test was performed. Down, downregulated; up, upregulated.

Extended Data Fig. 10 Relatively low expression of Rsm sRNAs during infections.

a. Schematics of the Gac-Rsm cascade in *P. aeruginosa*. Components that promote biofilm (green) and planktonic (blue) lifestyles are color-coded. **b**. Relative transcript abundance (TPM, transcripts per million) of *sicX* compared to all protein-coding sequences (CDS), *rsmYZ* sRNAs, and the remaining non-coding sRNAs in 202 transcriptomes examined in Fig. 2a. Transcriptomes are sorted from the highest to the lowest *sicX* TPM. **c**. A heatmap showing the mean TPM of CDS, sRNAs, *rsmYZ*, and *sicX* in *P. aeruginosa* transcriptomes. **d**. β-galactosidase assays demonstrating the expression of *sicX* is independent from GacS-GacA. P_{sicX}-*lacZ*, *lacZ* transcriptionally fused to *sicX* promoter. MrT7, MAR2xT7 transposon. $n = 4$ in all experiments. Error bars represent standard deviation from the mean. Significant difference (compared to WT) was analyzed with a Two-tailed Mann-Whitney test (n.s., not significant).

Supplementary information

Supplementary Information

This file contains Supplementary Fig. 1, Tables 1–3 and References.

Reporting Summary

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 8

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Cao, P., Fleming, D., Moustafa, D.A. *et al.* A *Pseudomonas aeruginosa* small RNA regulates chronic and acute infection. *Nature* **618**, 358–364 (2023). https://doi.org/10.1038/s41586-023-06111-7

- Received: 26 October 2022
- Accepted: 21 April 2023
- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06111-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

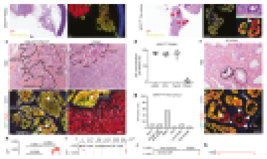
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

A bacterial small RNA controls the switch between chronic and acute infection

Nature Research Briefing 24 May 2023

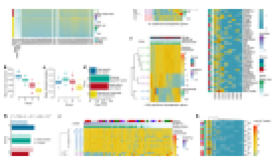
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06111-7



Ordered and deterministic cancer genome evolution after p53 loss

17 August 2022

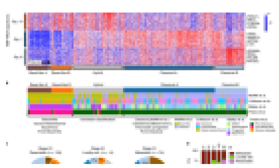
Timour Baslan, John P. Morris IV, … Scott W. Lowe



Phenotypic plasticity and genetic control in colorectal cancer evolution

26 October 2022

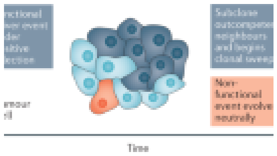
Jacob Househam, Timon Heide, … Trevor A. Graham



Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution

13 January 2020

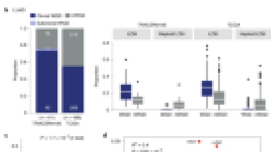
Michelle Chan-Seng-Yue, Jaeseung C. Kim, … Faiyaz Notta



[Genetic and non-genetic clonal diversity in cancer evolution]

16 March 2021

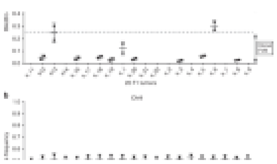
James R. M. Black & Nicholas McGranahan



[Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution]

05 March 2020

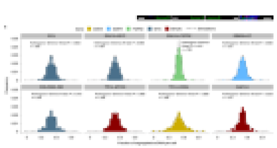
Saioa López, Emilia L. Lim, … Nicholas McGranahan



[Stochastic changes in gene expression promote chaotic dysregulation of homeostasis in clonal breast tumors]

14 June 2019

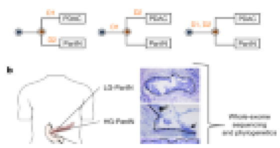
Sara J. Felts, Xiaojia Tang, … Larry R. Pease



The evolutionary dynamics of extrachromosomal DNA in human cancers

19 September 2022

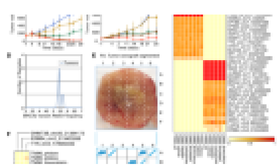
Joshua T. Lange, John C. Rose, … Paul S. Mischel



Precancerous neoplastic cells can move through the pancreatic ductal system

03 September 2018

Alvin P. Makohon-Moore, Karen Matsukuma, … Christine A. Iacobuzio-Donahue



High-resolution deconstruction of evolution induced by chemotherapy treatments in breast cancer xenografts

18 December 2018

Hyunsoo Kim, Pooja Kumar, … Jeffrey H. Chuang

- Article
- Open Access
- Published: 31 May 2023

Deterministic evolution and stringent selection during preneoplasia

- Kasper Karlsson,
- Moritz J. Przybilla,
- Eran Kotler,
- Aziz Khan,
- Hang Xu,
- Kremena Karagyozova,
- Alexandra Sockell,
- Wing H. Wong,
- Katherine Liu,
- Amanda Mah,
- Yuan-Hung Lo,
- Bingxin Lu,
- Kathleen E. Houlahan,
- Zhicheng Ma,
- Carlos J. Suarez,
- Chris P. Barnes,
- Calvin J. Kuo &
- …
- Christina Curtis

Show authors

Nature **volume 618**, pages 383–393 (2023)

- 10k Accesses

- 56 Altmetric

- Metrics details

Abstract

The earliest events during human tumour initiation, although poorly characterized, may hold clues to malignancy detection and prevention[1]. Here we model occult preneoplasia by biallelic inactivation of *TP53*, a common early event in gastric cancer, in human gastric organoids. Causal relationships between this initiating genetic lesion and resulting phenotypes were established using experimental evolution in multiple clonally derived cultures over 2 years. *TP53* loss elicited progressive aneuploidy, including copy number alterations and structural variants prevalent in gastric cancers, with evident preferred orders. Longitudinal single-cell sequencing of *TP53*-deficient gastric organoids similarly indicates progression towards malignant transcriptional programmes. Moreover, high-throughput lineage tracing with expressed cellular barcodes demonstrates reproducible dynamics whereby initially rare subclones with shared transcriptional programmes repeatedly attain clonal dominance. This powerful platform for experimental evolution exposes stringent selection, clonal interference and a marked degree of phenotypic convergence in premalignant epithelial organoids. These data imply predictability in the earliest stages of tumorigenesis and show evolutionary constraints and barriers to malignant transformation, with implications for earlier detection and interception of aggressive, genome-instable tumours.

Main

In rapidly adapting asexual populations, including microorganisms and tumours, multiple mutant lineages often compete for dominance[2]. These complex dynamics determine the outcomes of evolutionary adaptation but are difficult to observe in vivo. Experimental evolution has yielded fundamental insights into clonal dynamics in microorganisms, enabling characterization of mutant clones and their fitness benefits[3,4]. The same forces of mutation and selection fuel clonal expansions in somatic cells during ageing, contributing to malignancy, but their dynamics are poorly understood[5,6,7].

Cancers arise from a mutated cell that undergoes premalignant clonal expansion while accruing additional mutations. These mutations can spread in phenotypically normal tissues before apparent morphological changes, with aneuploidy and driver mutations preceding cancer diagnosis by years[5,8,9]. Identification of the causes of, and barriers to, malignant transformation requires characterization of the molecular phenotypes that precede this event in a tissue-specific manner. However, repeated sampling of healthy or preneoplastic tissue is impractical and thus evolutionary dynamics have been inferred from sequencing data[5,6,10]. For example, we inferred stringent subclonal selection in premalignant Barrett's oesophagus, whereas matched adenocarcinomas largely exhibited neutral evolution[6], presumably due to rapid growth after transformation and diminishing returns epistasis[11]. Despite these insights, the order of somatic alterations and patterns of clonal expansion that precede transformation are obscured in established cancers[5,12], necessitating new approaches to empirically measure premalignant evolution.

Gastric cancer (GC), the fourth-leading cause of cancer mortality worldwide, lacks routine screening albeit its long lead times contributing to late diagnoses, poor prognosis and limited treatment options[13,14]. Therefore, it is crucial to identify the molecular determinants of GC and its non-obligate precursor, intestinal metaplasia, which is poorly characterized compared with precursor lesions in the adjacent oesophagus (Barrett's oesophagus)[15,16,17]. Although the utility of forward-genetic and GC organoids as preclinical models has been established[18,19,20,21], in the former, combinatorial hits were engineered to bypass nascent progression and accelerate transformation[19,21].

Here we model tumorigenesis from the 'bottom up' using CRISPR–Cas9-engineered human gastric organoids (HGOs) to identify causal relationships between initiation of genetic insults and resultant genotypes and phenotypes. Because *TP53* inactivation is a common early event preceding numerical and structural chromosomal abnormalities (aneuploidy) in chromosomal instable (CIN) GC[19,22,23], we use non-malignant HGOs as a tabula rasa to study preneoplasia induced by *TP53* deficiency over a 2-year time span. HGOs are ideal for this task because they recapitulate the cellular attributes

of in vivo models, including three-dimensional tissue structure, multilineage differentiation and disease pathology[21].

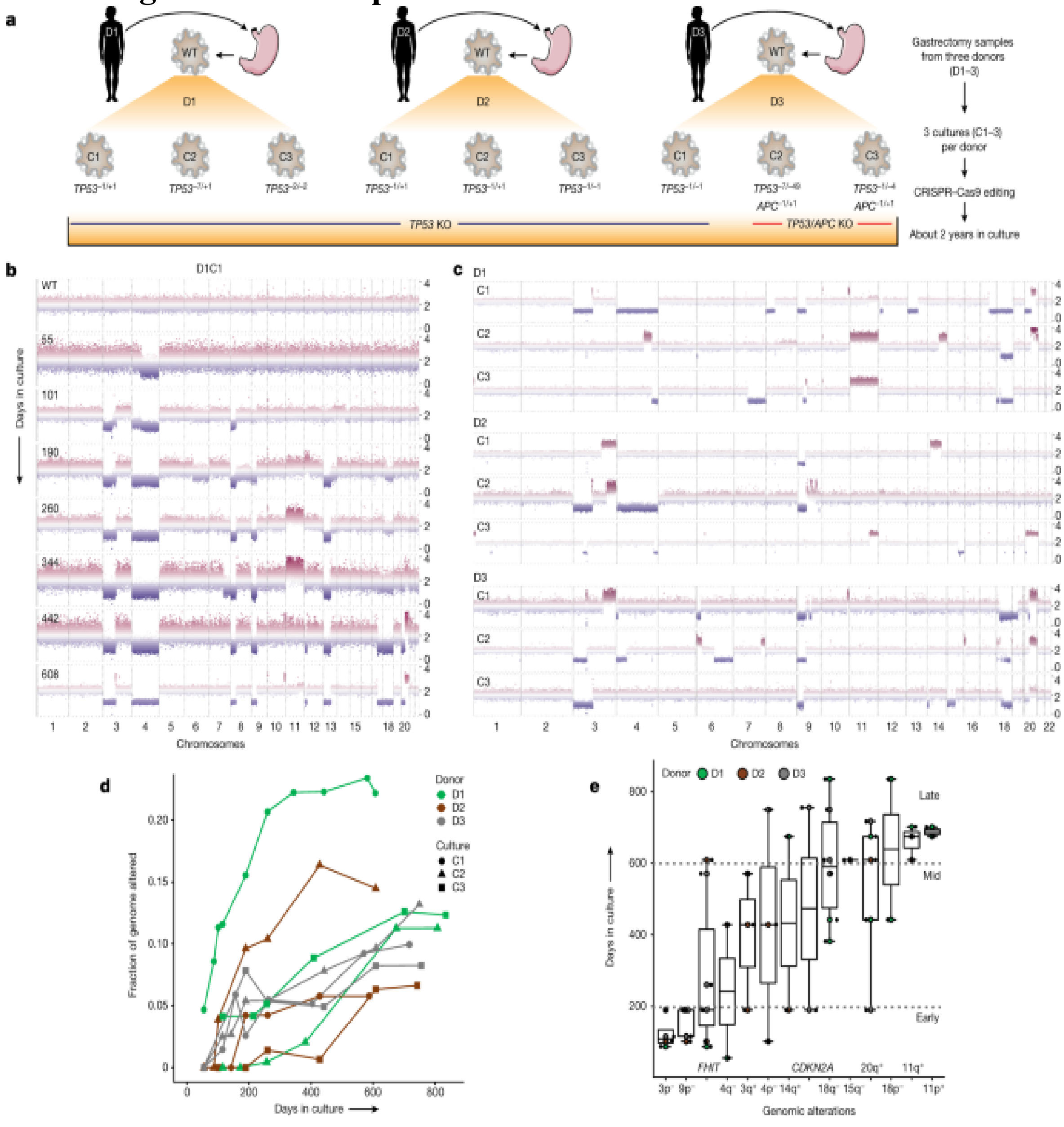
Whereas *TP53* is altered in over 70% of CIN GCs[22,23], its ability to elicit aneuploidy, a hallmark of most solid cancers, has been controversial and appears tissue dependent[24,25,26,27]. Moreover, the extent to which specific copy number alterations (CNAs) are selectively advantageous, and their tumorigenic impact is largely unknown[28,29]. We chart genotype-to-phenotype maps of gastric preneoplasia following *TP53* inactivation in multiple HGO cultures and demonstrate that these models recapitulate genomic hallmarks of gastro-oesophageal tumorigenesis, including the multi-hit, temporal and repeated acquisition of CNAs and structural variants (SVs), accompanied by progression towards malignant transcriptional states. Prospective lineage tracing with linked single-cell expression profiles delineates early clonal dynamics, showing extensive clonal interference, stringent selection and rapid adaptation, underpinned by temporal genomic contingencies and phenotypic convergence. Our findings highlight the power of experimental evolution in human organoids to investigate occult preneoplastic processes and the repeatability of somatic evolution.

TP53 $^{-/-}$ induces CNAs in defined orders

To model tumour initiation in CIN GC, we established HGOs from non-malignant tissue from three human donors undergoing gastrectomy and introduced biallelic *TP53* frameshift mutations via CRISPR–Cas9, resulting in an inactive gene product (Fig. 1a, Extended Data Fig. 1, Supplementary Figs. 1–3, Supplementary Tables 1 and 2 and Methods). From each donor (D1–3), three independent, clonally derived *TP53* $^{-/-}$ cultures (C1–3) were established, yielding nine cultures for long-term propagation, five of which were each split into three replicates (R1–3) for cellular barcoding studies ($n = 24$ cultures). Another ‘hit’ in the *APC* tumour suppressor, a Wnt pathway negative regulator altered in 20% of CIN GC (Extended Data Fig. 2a), was concurrently engineered in C2 and C3 from D3 (referred to as D3C2 and D3C3, respectively; Supplementary Figs. 1 and 3) to examine the evolutionary consequence of dual tumour suppressor inactivation. The clonal status of CRISPR-edited sites was verified via Sanger sequencing and

confirmed by whole-genome sequencing (WGS) at multiple time points (Supplementary Figs. 1–3). Throughout, we refer to time as days after *TP53* deficiency was engineered and we group *TP53*$^{-/-}$ and *TP53*$^{-/-}$ and *APC*$^{-/-}$ cultures unless otherwise specified.

Fig. 1: *TP53* deficiency in HGOs induces aneuploidy and GC-associated CNAs along a defined temporal order.



a, Schematic overview of HGO establishment and generation of *TP53*$^{-/-}$, *TP53*$^{-/-}$ and *APC*$^{-/-}$ cultures via CRISPR–Cas9 editing (Methods). **b**,

Genome-wide CNA profiles of D1C1 at multiple time points assessed by sWGS. Normalized read counts across 50 kb genomic windows for each time point. **c**, CNA profiles for the nine organoid cultures sampled between days 588 and 835. **d**, FGA over time for each culture. **e**, Time of appearance (days in culture) of persistent arm-level CNAs (or alterations in *FHIT* and *CDKN2A*) in *TP53*/*APC*-deficient organoids (alterations that became extinct were not considered). The prevalence of these alterations in GC is summarized in Extended Data Fig. 2a,b. Boxes show interquartile range (IQR), centre lines represent the median and whiskers extend by 1.5× IQR. Sample size per aberration (n): 3p^{-} (4), 9p^{-} (7), *FHIT* (7), 4q^{-} (2), 3q^{+} (3), 4p^{-} (3), 14q^{+} (2), *CDKN2A* (2), 18q^{-} (6), 15q^{-} (2), 20q^{+} (5), 18p^{-} (2), 11q^{+} (3), 11p^{+} (2). KO, knockout. Image of stomach in **a** is from Servier Medical Art, CC BY 3.0.

Source Data

We first asked whether *TP53* deficiency elicits aneuploidy, measured as numeric and/or structural chromosomal abnormalities (genome instability). To investigate CNAs we sequenced the nine cultures (shallow whole-genome sequencing, sWGS; median 0.2× coverage) at up to 11 time points, spanning early (0–200 days), mid (around 200–400 days) and late (around 400–900 days) intervals (Extended Data Fig. 1 and Supplementary Table 3). *TP53*-deficient organoids progressively acquired CNAs, first accruing chromosome arm-level losses followed by copy number gains (Fig. 1b). By contrast, wild-type (WT) gastric organoids remained genomically stable in the long term (13–26 passages; Supplementary Figs. 4–6). Cultures from the same donor exhibited variable CNA patterns, suggesting that genetic background does not wholly constrain subsequent alterations (Fig. 1c and Supplementary Figs. 4–6). Despite this variability, CNAs prevalent in the TCGA GC cohort were recurrently altered in *TP53*$^{-/-}$ cultures, including loss of chromosome (chr) 3p, 9p and 18q and gain of 20q (Extended Data Fig. 2)[22]. Additionally, arm-level CNAs present in two or more *TP53*$^{-/-}$ HGOs were enriched in gastric and oesophageal cancers but not in other tumour types ($P < 0.05$, two-sided Wilcoxon rank-sum test; Extended Data Fig. 2c). Thus, recurrent tissue-specific CNAs accrue in *TP53*$^{-/-}$ HGOs. During the experiment, because mycoplasma was detected in early-passage WT cultures and some derivative samples an antibiotic (normocin) was used

to eliminate infections (Methods and Supplementary Fig. 7a–f). Accordingly replicate experiments were performed under mycoplasma-free conditions, demonstrating that mycoplasma infection is not associated with CNAs or other molecular features (Extended Data Fig. 1, Methods and Supplementary Figs. 7g and 8).

Across all cultures the fraction of genome altered (FGA), a measure of aneuploidy, increased over time at varying rates and plateaued around day 600 (Fig. 1d and Methods). For example, D1C1, which accrued early arm-level alterations, exhibited over 20% FGA by day 260 compared with a median FGA of about 5% across all cultures at similar time points. *TP53*$^{-/-}$ and *TP53*$^{-/-}$/*APC*$^{-/-}$ cultures exhibited comparable FGA at final time points (average 11.3 and 10.7%, respectively), consistent with the expectation that *APC* loss does not fuel gastric cell aneuploidy. In several cultures FGA decreased over an interval due to clonal extinction (D3C3 day 190 versus day 442; D2C2 day 428 versus day 609) (Fig. 1d and Supplementary Figs. 5 and 6). As expected, FGA was lower in *TP53*$^{-/-}$ HGOs than in CIN GCs (median FGA 34.5% in TCGA, according to cBioPortal).

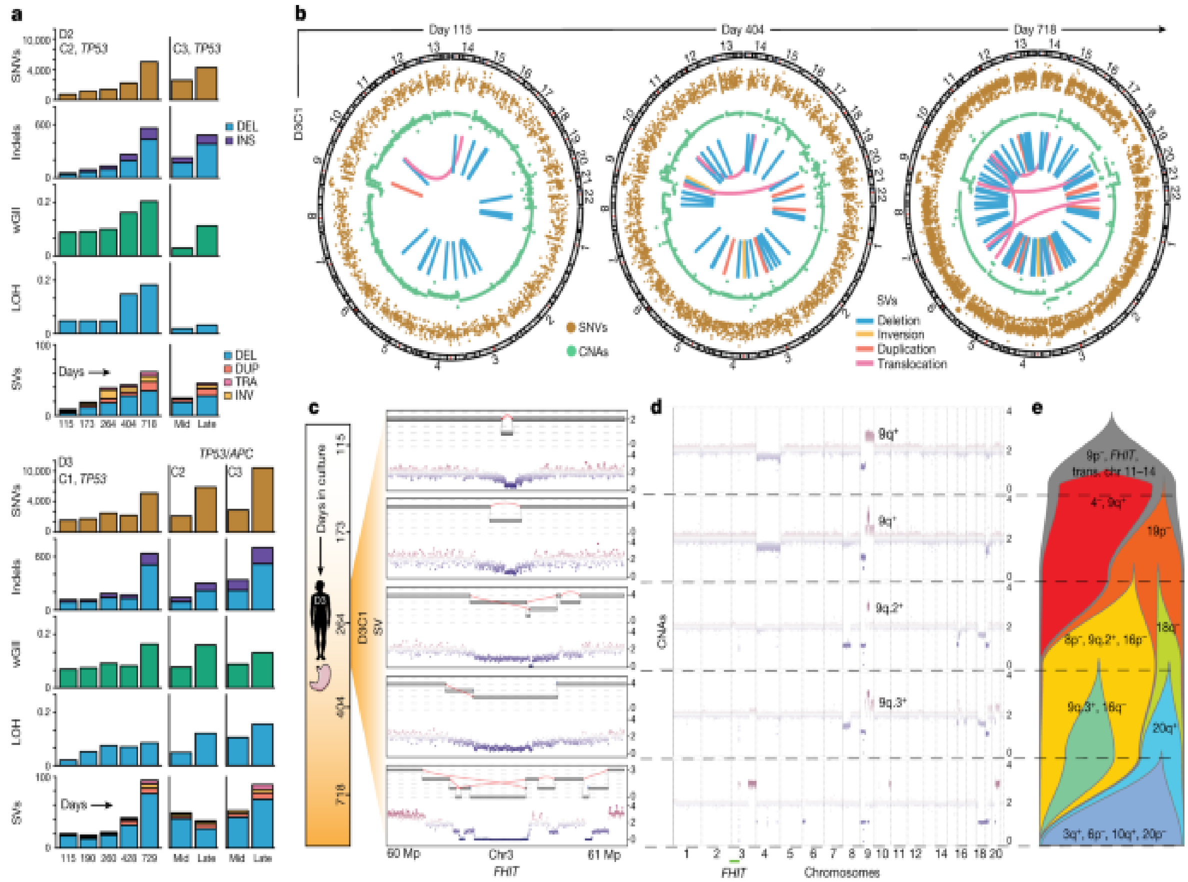
Investigation of the temporal onset of arm-level and focal CNAs in *TP53*$^{-/-}$ HGOs showed preferred orders (Fig. 1e and Supplementary Table 3). Specifically, loss of chr9p and chr3p repeatedly occurred (across donors and cultures) within 200 days but seldom later, suggesting a period during which these alterations were particularly advantageous. Chr9p deletion spans the *CDKN2A* tumour suppressor commonly altered in the CIN subgroup of gastric (roughly 41%; Extended Data Fig. 2a) and oesophageal (roughly 74%) adenocarcinomas, and co-occurs with *TP53* alterations[19,22]. Indeed, *CDKN2A* loss signals the initiation of Barrett's oesophagus progression to dysplasia and oesophageal adenocarcinoma[30] and GC premalignancy[19]. Deeper sequencing confirmed biallelic loss of *CDKN2A* via focal deletion (D3C1, D3C3) or truncating mutations in p16 (*INK4A*), along with heterozygous loss (D1C3) (Supplementary Fig. 9 and Supplementary Table 3). Similarly, deletion of the FHIT/FRA3B protein encoded on chr3p commonly occurred early in *TP53*$^{-/-}$ HGOs (median 190 days) (Fig. 1e, Supplementary Fig. 10 and Supplementary Table 3). A genome caretaker, *FHIT*, is lost early during tumour progression leading to deoxythymidine triphosphate depletion, replication stress and DNA breaks[31]. Notably, 12%

of CIN GCs harbour *FHIT* alterations (Extended Data Fig. 2a). Although *CDKN2A* and *FHIT* deletions are insufficient for malignant transformation[15,32], their recurrent early loss during in vitro evolution and in GCs implies a role in tumour initiation. Additional GC-associated CNAs include loss of chr18q and gain of chr20q, which consistently occurred late (around 600 days). Such late alterations may reflect dynamic selective pressures from increased fitness or new evolutionary paths enabled by earlier alterations[4]. These data demonstrate that *TP53* loss facilitates aneuploidy in gastric cells and accrual of tissue-specific CNAs in a defined order.

Selection and clonal interference

We next sequenced (WGS, mean coverage 26×) five *TP53*$^{-/-}$ cultures at multiple time points (Fig. 2a, Extended Data Fig. 3a and Supplementary Table 4). This confirmed biallelic *TP53* and *APC* inactivation at CRISPR target sites (Supplementary Fig. 2) and showed an increase in the weighted genome instability index (wGII), the fraction of genome with loss of heterozygosity (LOH), as well as focal deletions and amplifications during prolonged culture (Fig. 2a, Extended Data Fig. 3b and Supplementary Table 4). Single-nucleotide variants (SNVs) and SVs also increased over time (Fig. 2a, Extended Data Fig. 3b and Supplementary Table 4). At late time points the SNV burden was higher in *TP53*$^{-/-}$/*AP*C$^{-/-}$ (D3C2, D3C3) than in *TP53*$^{-/-}$ HGOs. Few GC-associated genes were mutated across donors (Extended Data Fig. 2e).

Fig. 2: *TP53* deficiency elicits subclonal copy number evolution, SVs and clonal interference.



a, Burden of different classes of somatic genomic alterations in *TP53*$^{-/-}$ and *TP53*$^{-/-}$, *APC*$^{-/-}$ HGOs (relative to WT over time), as assessed by longitudinal WGS of individual cultures at the specified time points (mid, day 296; late, days 743–756). **b**, Circos plots for D3C1 illustrating increasing genomic instability and complexity over time. Classes of alterations shown include SNVs (adjusted variant allele frequencies), CNAs (log(*R*)) and SV consensus calls ([Methods](#)). **c**, Evolution of rigma-like SVs at the *FHIT* fragile site on chr3p. Zoomed-in view of a 1 Mb region in the *FHIT* locus. Top, reconstructed SVs; bottom, corresponding CNAs. **d**, Longitudinal CNA profiles for D3C1. **e**, Fishplot schematic for D3C1 illustrating subclonal CNA evolution, clonal interference and extinction. Subclone frequencies (*x* axis) were determined based on CNAs visualized in **d** ([Methods](#)). DEL, deletion; DUP, duplication; INV, inversion; TRA, translocation. Image of stomach in **c** is from Servier Medical Art, [CC BY 3.0](#).

[Source Data](#)

Several regions of densely clustered mutations (hypermutation) were noted, including the *FHIT* fragile site in all sequenced cultures (Supplementary Figs. 11–13 and Supplementary Table 3). WT cultures exhibited simple focal *FHIT* deletions at late passages, probably due to clonal expansion of an initially rare event, and suggestive of somatic mosaicism (Supplementary Figs. 11a,b and 12a,b). Single-base substitutions 1, 5 and 40, which are ubiquitous and implicated in ageing and cancer, were the most prevalent mutational signatures[33]. However, by the late time point D3C2 developed single-base substitution 17a/b (Extended Data Fig. 3b and Supplementary Table 4), which is prevalent in gastro-oesophageal carcinomas and progressive Barrett's oesophagus lesions[32].

All classes of alterations accumulated in evolved *TP53*$^{-/-}$ HGOs but SVs were particularly notable, with non-clustered and simple clustered rearrangements dominating at early time points followed by complex clusters (ten or more rearrangements) involving deletions, inversions and translocations over time (Fig. 2a and Extended Data Fig. 3c,d). This is exemplified by D3C1, which accrued multiple interchromosomal rearrangements (Fig. 2b). Although such complex SVs are seldom reported in normal tissues, they are prevalent in progressive Barrett's oesophagus[16]. SV burden increased markedly in *TP53*$^{-/-}$ HGOs between early and late time points (median change of 148%), exceeding by over threefold the change in SV burden (45%) between endoscopies in patients with Barrett's oesophagus harbouring biallelic *TP53* inactivation and who subsequently progressed to oesophageal adenocarcinoma (average 2.2 years, range 0.65–6.16 years)[16]. By contrast, Barrett's oesophagus non-progressors (lacking *TP53* biallelic inactivation) had a low and stable SV burden between endoscopies (Extended Data Fig. 3e).

The *FHIT* locus frequently harboured complex SVs, including deletion chasms at fragile sites (rigma), as reported in GC and Barrett's oesophagus[34] (Fig. 2c, Extended Data Fig. 4a and Methods). We traced the genesis of rearrangements at the *FHIT* in D3C1, starting from a small deletion at day 115 and culminating in rigma by day 264. The subclone harbouring this rearrangement was lost (Fig. 2d,e, yellow subclone) but a separate subclone (blue) with a distinct *FHIT* rigma emerged and persisted, suggesting convergent evolution. Thus, rearrangements with multiple junctions evolve

over several generations, not as a single event as previously proposed[35]. Similar events evolved in other cultures, including a chr3 and chr9 translocation (Extended Data Fig. 4b–f and Supplementary Fig. 14a,b). Despite these rearrangements, overall genomic content remained diploid as confirmed by flow cytometry (Supplementary Fig. 14c and Supplementary Table 4).

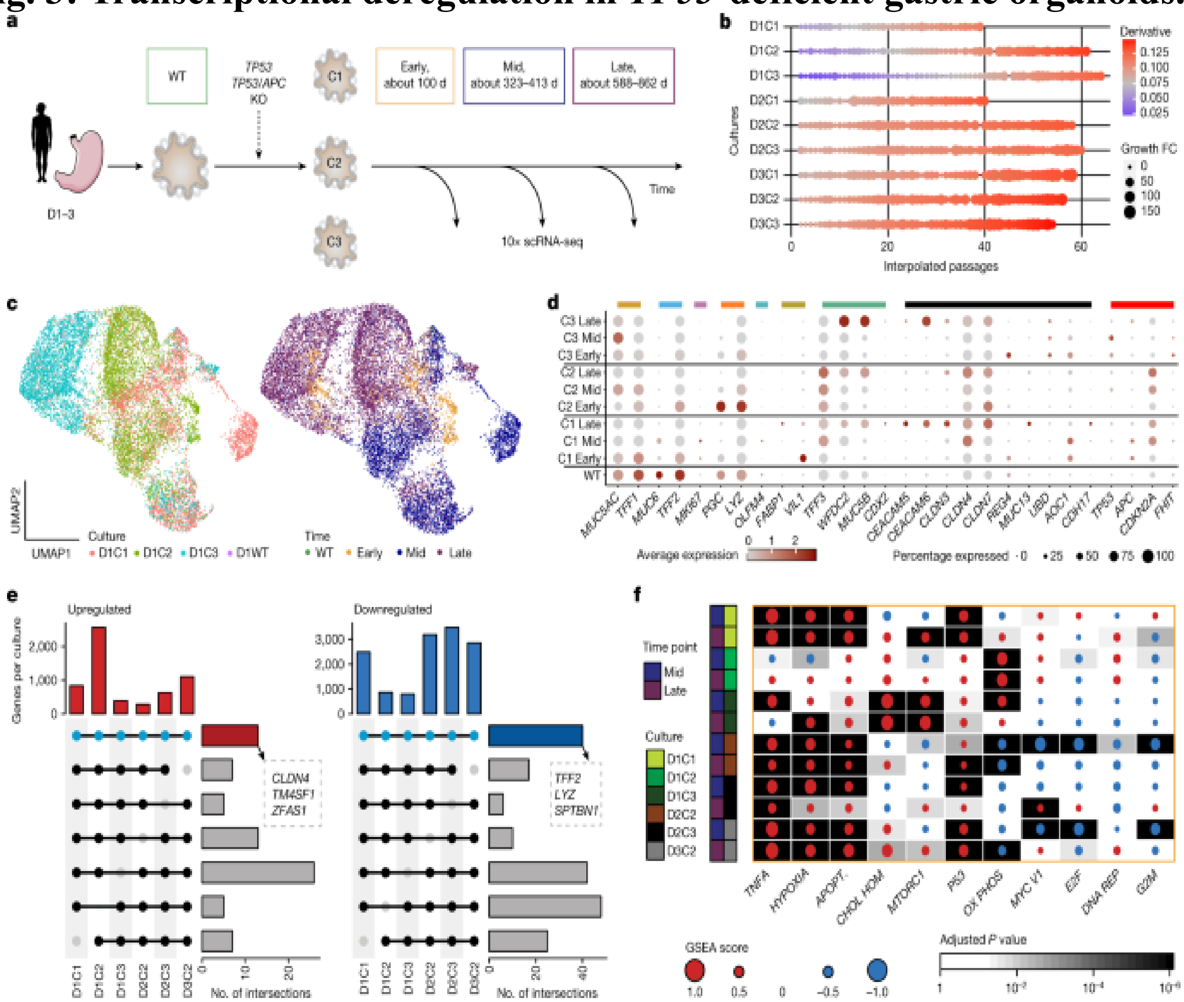
Clonal competition and extinction were investigated by determination of subclonal populations from CNA profiles (via bulk WGS) across five time points for D3C1 and D2C2. By day 115, D3C1 had acquired numerous deletions (9p, *FHIT*) and several SVs, including a persistent chr11–chr14 translocation. Over 600 days, multiple CNA-defined subclones increased in frequency before extinction (Fig. 2d,e and Supplementary Table 5). For example, a chr4^{-}, 9q^{+} subclone arose early but disappeared by day 264, outcompeted by a chr19p^{-} subclone that later acquired chr8p^{-}, 9q.2^{+} and 16p^{-} alterations and remained dominant until day 404. This subclone was ultimately outcompeted by one with chr18q loss that acquired gain of chr20q, both recurrent late events in multiple cultures. Thus, some clones fix and achieve dominance whereas others reach substantial frequencies before going extinct, presumably due to clonal interference. Distinct CNA subclones coexisted for extended durations (around 140 days), suggesting comparable fitness (for example, chr8p^{-}, 9q.2^{+}, 16p^{-} and 18q^{-} subclones) and intermittent periods of clonal competition and stasis as seen in other cultures (Extended Data Fig. 4c,d). These data demonstrate stringent selection and pervasive clonal interference in premalignant epithelial populations.

Transcriptional changes following *TP53* $^{-/-}$

Phenotypic and transcriptional changes during in vitro evolution were evaluated based on growth dynamics of TP53$^{-/-}$ HGO cultures and single-cell RNA sequencing (scRNA-seq) at early, mid and late time points (Fig. 3a, Extended Data Figs. 1 and 3a, Supplementary Table 2 and Methods). We investigated changes in cell proliferation by fitting a Loess regression model to cell numbers at each passage, using growth derivative and fold change as a surrogate for fitness. Higher growth derivatives were observed at late and

mid versus early time points (Fig. 3b and Supplementary Fig. 15a); the use of raw cell numbers yielded similar results ($P = 0.003$, two-way repeated-measures analysis of variance; Supplementary Fig. 15b). scRNA-seq of 12 cultures (seven *TP53*$^{-/-}$, two *APC*/*TP53* and three WT) yielded 31,606 cells for analysis following quality control (Fig. 3c, Extended Data Fig. 5a,b, Supplementary Fig. 16, Supplementary Table 6 and Methods). Normal gastric tissue markers were expressed in WT HGOs from the three donors, including pit mucosal cell (PMC) markers (*MUC5AC*, *TFF1*, *TFF2*, *GKN2*) in D1, enterocyte markers (*FABP1*, *FABP2*, *ANPEP*, *PHGR1*, *KRT20*) in D2 and gland mucosal cell (GMC) markers (*MUC6*, *PGC*, *TFF2*, *LYZ*) in D3, but were heterogeneous in *TP53*$^{-/-}$ HGOs (Fig. 3c–e, Extended Data Fig. 5, Supplementary Fig. 16 and Supplementary Table 6).

Fig. 3: Transcriptional deregulation in *TP53*-deficient gastric organoids.



a, Experimental overview of longitudinal scRNA-seq profiling of gastric organoid cultures. WT and replicate *TP53*$^{-/-}$ HGOs were sampled at multiple time points (early, about 100 days, orange; mid, about 320 days, blue; late, about 770 days, purple) and subjected to scRNA-seq. **b**, Dot-plot depicting estimated growth curve derivatives and growth fold change (FC) from previous time points for each culture over time (interpolated passage number). **c**, Uniform manifold approximation and projection (UMAP) visualizations coloured according to culture (left) and time point (right) for D1, depicting 13,984 cells. **d**, Dot-plot depicting the expression of selected marker genes for individual cultures and time points. Coloured bars highlight (1) marker genes associated with normal gastric and intestinal cell types, (2) genes upregulated in gene expression profiling interactive analysis (GEPIA) of GC and (3) others of functional relevance. PMCs, *MUC5AC*, *TFF1*, dark yellow; GMCs, *MUC6*, *TFF2*, light blue; proliferative cells, *MKI67*, purple; neck-like cells, *PGC*, *LYZ*, orange; mucosal stem cells, *OLFM4*, turquoise; enterocytes, *FABP1*, *VIL1*, olive; goblet cells, *TFF3*, *WFDC2*, *MUC5B*, *CDX2*, green; GEPIA top 12 genes, *CEACAM5*, *CEACAM6*, *CLDN3*, *CLDN4*, *CLDN7*, *REG4*, *MUC3A*, *MUC13*, *PI3*, *UBD*, *AOC1*, *CDH17*, black; other, *TP53*, *APC*, *CDKN2A*, *FHIT*, red. **e**, UpSet plot representing shared differentially up- (left) and downregulated genes (right) across donors and cultures ($P < 0.05$, Bonferroni corrected two-sided Wilcoxon rank-sum test). **f**, GSEA heatmap for MsigDB Hallmark gene sets showing pathways most significantly altered for each culture (Kolmogorov–Smirnov statistic, Benjamini–Hochberg adjusted, two-sided). GSEA score is indicated (dot size) and coloured according to the directionality of expression profiles (up, red; down, blue). Image of stomach in **a** is from Servier Medical Art, CC BY 3.0.

Source Data

The mucosal-like phenotype in WT cultures, defined by mucin and TFF gene expression, was lost following *TP53*$^{-/-}$ in D1 and D3. Additionally, in D1 intestinal goblet cell-specific markers—including *TFF3*, *WFDC2* and *MUC5B*—were upregulated at the late time point, as commonly seen in intestinal metaplasia[36]. GC-associated genes, including claudins (*CLDN3*, *CLDN4*, *CLDN7*) and the carcinoembryonic antigen (CEA) family (*CEACAM5*, *CEACAM6*) increased in expression over time in D1 and D3.

The inverse was observed in D2 cultures, plausibly due to an inflamed biopsy and the predominance of enterocytes in WT culture[14] (Extended Data Fig. 5a,c). The absence of *MUC5AC* following *TP53*$^{-/-}$ and increase in *CEACAM6* expression was verified by immunofluorescence staining in D3C2 (Supplementary Fig. 15c).

We investigated the overlap in transcriptional features across *TP53*$^{-/-}$ HGOs by intersection of significantly differentially expressed genes (DEGs) from early to late time points across the six cultures with scRNA-seq data. In total, 13 consistently upregulated and 40 downregulated genes were identified (Bonferroni corrected $P < 0.05$, Wilcoxon rank-sum test; Fig. 3e and Supplementary Table 6). Upregulated genes included *CLDN4*, *TM4SF1* and *ZFAS1*, which are implicated in GC[22,37], whereas those downregulated included *SPTBN1*, a cytoskeletal protein involved in TGFβ signalling[38], and mucin production modulators *LYZ* and *TFF2*.

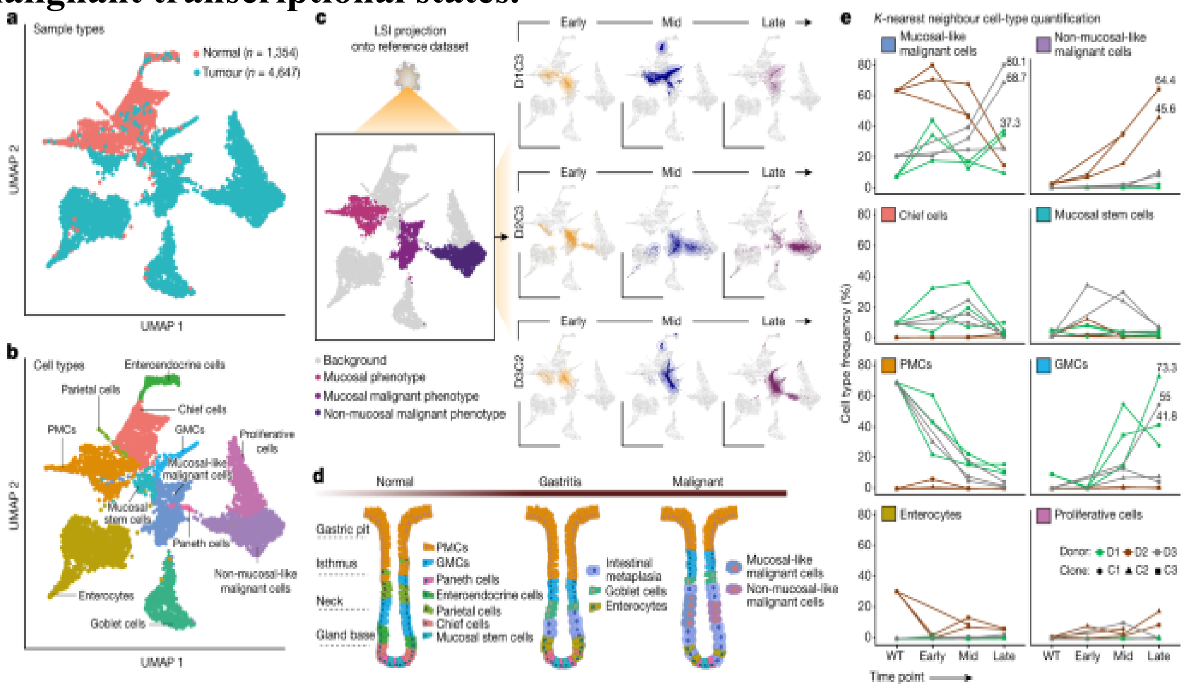
Last, we assessed pathway-level changes by gene set enrichment analysis (GSEA) of DEGs in both late versus early and mid versus early time points (Fig. 3f and Supplementary Table 6). Several pathways were enriched across multiple cultures and donors, including upregulation of tumour necrosis factor (TNF) signalling via nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), as reported in CIN tumours[39] and comparisons of GC versus normal tissue[40] (four of six cultures), apoptosis (five of six cultures) and hypoxia (five of six cultures). Downregulated pathways included MYC, E2F targets and G2M checkpoints, although these were more variable and probably reflect survival programmes. Thus, despite heterogenous single-gene trajectories, pathways implicated in malignancy were shared across cultures and donors.

Emergence of malignant expression states

To identify pathologic features we projected HGO longitudinal scRNA-seq data onto a reference atlas comprising both normal and GC scRNA-seq[41] (Methods). Restriction of the reference to epithelial cells yielded 6,001 cells (1,354 normal, 4,647 tumour) assigned to distinct cell type clusters using literature-derived marker genes[40,41,42]. Two tumour cell clusters emerged,

comprising mucosal-like malignant and non-mucosal-like malignant cells[41], the latter including malignant markers (*KRT17*, *KRT7*, *LY6D*) but lacking mucosal markers (*MUC5AC*, *TFF2*, *TFF1*) (Fig. 4a). PMCs, GMCs, chief cells, parietal cells, enterocytes, enteroendocrine cells, goblet cells and proliferative cells were also assigned to clusters (Fig. 4b).

Fig. 4: Unsupervised assessment indicates progression towards malignant transcriptional states.



a,**b**, UMAP embedding of 6,001 epithelial cells from the Sathe et al. gastric tumour–normal scRNA-seq dataset[41], coloured according to histology (**a**) and assigned cell type (**b**). Detected cell types included PMCs, GMCs, chief cells, parietal cells, enterocytes, enteroendocrine cells, goblet cells and proliferative cells, as well as two types of malignant cell (mucosal-like and non-mucosal-like). **c**, LSI projection of *TP53*$^{-/-}$ HGOs sampled at early (orange), mid (blue) and late (purple) time points onto the reference dataset (left), coloured by cellular phenotypes of interest, providing orientation for the LSI projection of the three HGO cultures at the specified time points (right). The density of projected cells is highlighted using two-dimensional density distribution. LSI, latent semantic index. **d**, Schematic representation of shifts in cell populations proposed to accompany the transition from normal tissue to gastritis and that can lead to intestinal metaplasia and malignancy, adapted from ref. [40]. **e**, Projected cell type frequencies based on

the 25 nearest neighbours in HGOs over time. Panel **d** created with BioRender.com.

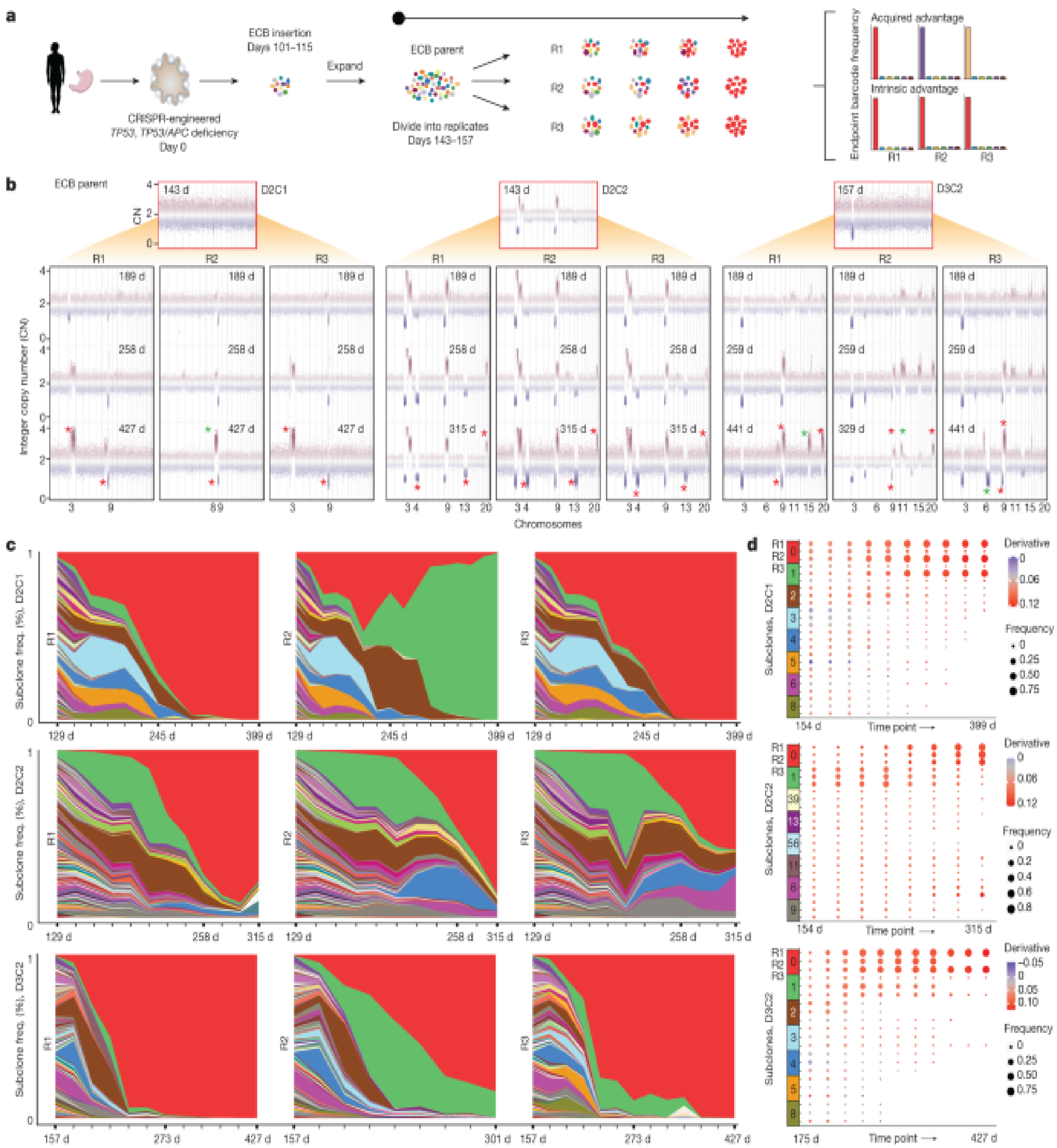
We next projected batch-corrected scRNA-seq data from early, mid and late time points individually onto the reference embedding to identify gastric cell types that were most similar (Fig. 4c, Extended Data Fig. 5e–h and Methods). Because the reference atlas lacked preneoplastic populations, cells were projected onto either normal or tumour cell state in which the majority of HGO cells mapped onto the latter. Shifts in cell states over time were evident for all cultures, some of which have been implicated in the normal-to-gastritis transition that can lead to intestinal metaplasia and ultimately malignancy (shown schematically in Fig. 4d). Changes in cell type frequencies were quantified for each HGO culture by identification of the 25 nearest neighbours (NNs) in the reference population (Fig. 4e). An increase in mucosal-like malignant cells was observed in three of seven cultures at the late time point, with 68.7, 80.1 and 37.3% of NNs being mucosal-like malignant cells for D3C2, D3C3 and D1C1, respectively. By contrast, for D2, mucosal-like malignant cells decreased whereas non-mucosal-like malignant cells increased from WT to the late time point (D2C2, 45.6%; D2C3, 64.4%, NNs) (Fig. 4e), explaining transcriptional differences relative to D1 and D3 (Fig. 3). Notably, approximately 30% of cells in D2WT projected near enterocytes, potentially contributing to gastritis-like features and underlining the transcriptional similarity between enterocytes and malignant cells[42]. WT cultures from D1 and D3 exhibited predominantly mucosal phenotypes. The decrease in mucosal gene expression suggests that the evolved *TP53*-deficient HGOs were en route towards intestinal metaplasia and malignancy, albeit at different rates, corroborating the supervised analyses based on specific marker genes. Although our HGO cultures harbour hallmarks of CIN GC, they do not exhibit evidence of histologic transformation (Supplementary Fig. 15d).

Deterministic growth of rare subclones

We next leveraged our HGO models to characterize preneoplastic subclonal dynamics at cellular resolution via prospective lineage tracing with high-complexity cellular barcodes. To jointly recover lineage and transcriptional states we developed expressed cellular barcodes (ECBs), which uniquely

label each cell (Supplementary Fig. 17a,b and Methods). Five *TP53*$^{-/-}$ (D1C1, D1C3, D2C1, D2C2, D2C3) and one *TP53*$^{-/-}$, *APC*$^{-/-}$ (D3C2) culture were transduced with ECB lentivirus between days 101 and 115 and evolved in parallel to the non-barcoded cultures for over 1 year. All cultures were Sanger sequenced at multiple time points to verify *TP53*/*APC* deletion clonality, resulting in the exclusion of D1C3 (Supplementary Fig. 17c,d). Each ECB parental line was split into three replicates to evaluate the reproducibility of clonal dynamics, in which outgrowth of the same subclone is assumed to reflect an intrinsic fitness advantage and divergent subclone dominance suggests acquired fitness differences (Fig. 5a).

Fig. 5: Lineage tracing showing subclonal dynamics and deterministic outgrowth.



a, Overview of prospective lineage tracing studies in *TP53*$^{-/-}$ HGOs using ECBs. The ECB parental population was split into replicates, and individual cultures evolved in parallel and subject to longitudinal barcode sequencing, showing subclonal dynamics and assessment of intrinsic or acquired fitness advantages amongst replicate cultures. **b**, CNA profiles were assessed by sWGS before the introduction of the ECB in the parental line and across replicate ECB cultures at multiple time points. Red asterisks denote CNAs present in at least two replicates but not in the parental population; green asterisks denote CNAs unique to one replicate. Only chromosomes

harbouring newly arisen CNAs (not present in the parental population) are numbered, for simplicity. **c**, Muller plots depicting ECB frequencies (assessed by barcode sequencing) over time, where each colour represents a distinct subclone in each replicate. Note that, for D3C2, R2 (D3C2R2), the barcode was lost around day 273. **d**, Dot-plots indicating ECB subclone frequency (indicated by size) and estimated growth curve derivative per subclone (indicated by colour). Image of stomach in **a** is from Servier Medical Art, CC BY 3.0.

Source Data

Longitudinal sWGS of these long-term ECB cultures demonstrated marked reproducibility at the genomic level, with recurrent CNAs shared across replicate cultures (Fig. 5b, Extended Data Fig. 6 and Supplementary Fig. 18). For example, in D2C2 new CNAs emerged around day 258 (loss of chr4q and chr13, gain of chr20q) across all three replicates. In D2C1, R2 (D2C1R2) gain of chr8q was detected by day 258 and persisted (Fig. 5b) but was mutually exclusive with gains of chr3q in R1 and R3. By contrast, CNAs in different cultures from the same donor were more variable (Fig. 1c).

Through DNA sequencing of ECBs at regular intervals, we estimated the relative abundances of subclones over time and constructed Muller plots to visualize clonal dynamics (Fig. 5c). Colours were assigned to barcodes based on subclone frequencies across replicates within a culture, with the highest-frequency subclone coloured red. For example, the red band in D2C1R1 represents the same barcoded subclone as in D2C1R2 and D2C1R3. For each culture (except D2C1R2) the same (red) subclone became dominant across all replicates (Fig. 5c), consistent with an intrinsic fitness advantage and deterministic outgrowth (Fig. 5a). For D2C1 replicates R1 and R3 the red subclone became dominant in line with their shared CNA profiles whereas in R2 the green subclone, which acquired a chr8q gain (spanning the *MYC* oncogene), overtook the population. Intriguingly, brown and green clones expanded concomitantly before going extinct, suggesting their mutual dependence.

Of note, subclone frequency correlations over time across replicates was generally high, reflecting similar subclonal dynamics within a culture and

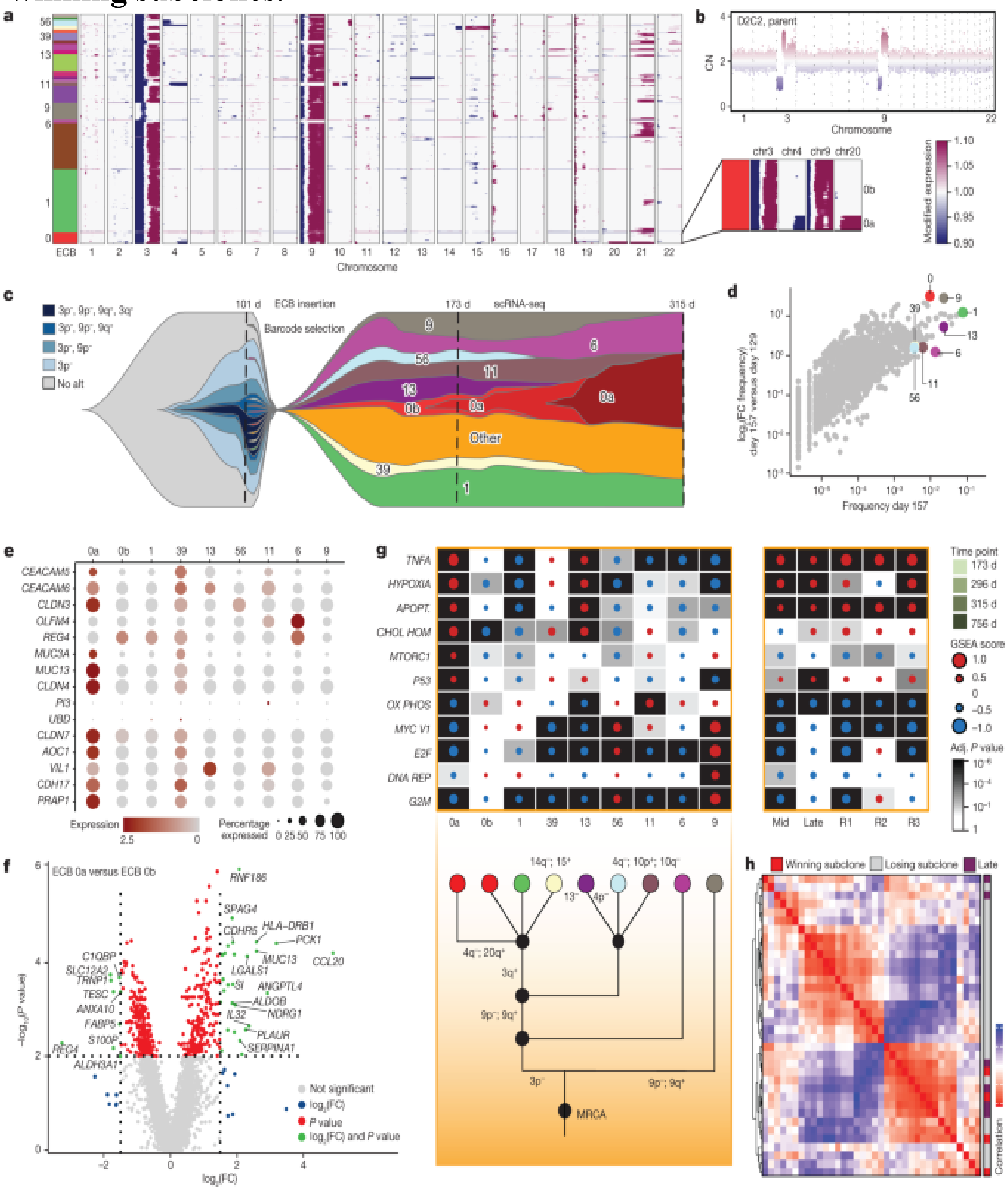
similar patterns across cultures (Extended Data Fig. 6). Especially striking was the remergence of the blue and purple subclones in D2C2R2 and D2C2R3 at around 200 days (Fig. 5c). By construction of subclone-specific growth curves and estimation of their derivatives, we found that ‘winning’ subclones had high initial fitness and increased in proliferative capacity over time (Fig. 5d and Methods). Thus lineage tracing shows reproducible dynamics across replicate cultures, with adaptive lineages sweeping rapidly to fixation and dominant clones comprising 75% (median across cultures) of the population by day 144 post ECB transduction (Supplementary Table 7). These patterns are reminiscent of rapid adaptation in isogenic microbial populations attributable to standing variation in the initial population[43,44].

Molecular features of winning subclones

To investigate the targets of selection and how they change over time and across populations, we leveraged ECBs jointly capturing lineage and transcriptional states in individual cells. Specifically, we sought to characterize the molecular features of winning subclones that dominated the population after prolonged evolution by performing scRNA-seq for several donors and replicates at selected time points when the population was heterogeneous. For D2C2R2, which was sampled at day 173, 1,284 cells passed quality control and we identified 20 subclones with at least ten cells, all of which were among the top 38 most frequent ECBs based on barcode sequencing. Arm-level CNAs were inferred from the scRNA-seq data using *inferCNV* (Methods), showing numerous subclone-specific CNAs (Figs. 5b and 6a and Methods). Reassuringly, aggregate CNA landscapes were concordant with WGS data and scDNA-seq showed profiles and frequencies similar to subclone-specific CNAs inferred from scRNA-seq (Supplementary Fig. 19 and Methods). A detailed examination of this replicate (D2C2R2) showed complex evolutionary dynamics amongst coexisting subclones. Most cells comprising the winning subclone (ECB-0, red) acquired chromosome $3p^{-}$, $3q^{+}$, $9p^{-}$ and $9q^{+}$ alterations early because these events were clonal or nearly clonal in the parent population at day 143 (Fig. 6a,b). A subpopulation within ECB-0 (termed 0a) additionally acquired $chr4q^{-}$ and $chr20q^{+}$ and ultimately became dominant, with these alterations present in roughly 90% of the population at day 315 (Fig. 5b and Supplementary Fig. 20). Similar dynamics were seen across all replicate

cultures in which winning subclones contained a nested CNA-defined subclone (Extended Data Figs. 7–9). These patterns may reflect a 'rich-get-richer' effect whereas fitness advantages acquired early drive clonal expansions, thereby increasing the likelihood of additional alterations that fuel growth[45].

Fig. 6: Genotype-to-phenotype mapping defines molecular determinants of winning subclones.



a, Inferred CNA heatmap from scRNA-seq data for D2C2R2 at day 173, where each row represents a cell. Colour bar at the left indicates the ECB to which each cell maps. Numbered barcodes were selected for further investigation. Inset shows a subpopulation within ECB-0 with additional CNAs, termed 0a, and the ECB-0 parent subclone is termed 0b. **b**, CNA profile for the D2C2 parental population (also shown in Fig. 5b). **c**, Fishplot schematic illustrating the link between lineage (ECBs) and CNA subclones. To facilitate visualization, subclones of interest (denoted in **a**) are shown and the remainder grouped as 'other'; all values are log transformed. **d**, Scatterplot comparing subclone frequency at day 157 and $\log_2$(FC) between days 129 and 157. All subclones are shown, with those of interest highlighted as in **a**. **e**, Dot-plot showing the expression of top DEGs based on GEPIA of gastric cancers. **f**, Volcano plot illustrating DEGs from comparison of the winning subclone 0a and its parental subclone 0b. Vertical and horizontal lines correspond to absolute $\log_2$(FC) values of 1.5 and $P < 0.01$ (two-sided Wilcoxon rank-sum test, not corrected for multiple testing), respectively. **g**, GSEA heatmap from MsigDB Hallmark gene sets showing the most significantly altered pathways (Kolmogorov–Smirnov statistic, Benjamini–Hochberg adjusted, two-sided) for specific subclones at day 173 (left) and later time points for the same culture (right). A manually reconstructed phylogeny is shown below. **h**, Pairwise Spearman correlation between samples based on GSEA score for the top ten most altered pathways for late relative to early time points and for subclones from multiple ECB replicate experiments. MRCA, most recent common ancestor.

Source Data

Because successful subclones consistently acquired additional genetic diversity, we sought to investigate the functional relevance of these events, focusing on a subset of subclones with divergent CNAs. As an example, D2C2R2 consisted of at least five different CNA clones at the time of barcode insertion (Fig. 6c and Supplementary Table 8). Multiple instances of convergent evolution were evident within this culture, in which subclones acquired the same CNA independently, implying stringent selection. For example, ECB-0a, ECB-11 and ECB-56 each lost variable-sized regions of chr4q. ECB-9 lacked common early alterations including chr3p^{-}, but subsequently acquired chr9p/q alterations. Despite the incomplete set of

CNAs, the growth of ECB-9 closely trailed that of the winning subclone (ECB-0; Fig. 6d). Convergent CNA evolution was also evident across cultures, in which chr15 and chr20 amplifications were present in the majority of cells in D3C2R1 at day 441, and these events plus chr11 amplification were present in R2 and R3 subclones by day 259.

Although highly fit subclones differed in genomic landscapes, we reasoned that they would share transcriptional programmes. Indeed D2C2R2, the winning subclone 0a (but not its parent, 0b) exhibited high expression of several GC genes including *CEACAM5*, *CEACAM6*, *CLDN3*, *CLDN4* and *CLDN7* (Fig. 6e). These genes were also highly expressed in winning subclones of all other replicate cultures, except for ECB-1a (green) in D2C1R2, which acquired 8q gain (Extended Data Figs. 7e, 8d and 9d). The winning subclone, 0a (versus 0b), also upregulated GC genes, including *RNF186* which regulates intestinal homeostasis and is associated with ulcerative colitis[46]; *MUC13*, which encodes a transmembrane mucin glycoprotein[47]; *CCL20*, a chemokine and candidate biomarker[48]; and *LGALS1* (galectin-1), which promotes epithelial–mesenchymal transition, invasion and vascular mimicry[49] (Fig. 6f and Supplementary Table 9). GSEA analysis, comparing the winning subclone in D2C2 with all other cells, showed upregulation of several pathways including TNF signalling via NF-κB, as well as hypoxia, apoptosis and p53 (Fig. 6g, Extended Data Fig. 10a and Supplementary Table 9). These same pathways were upregulated in three barcoded replicates for D2C2 at the final time point (day 315), as well as in the non-barcoded D2C2 culture at mid and late time points (relative to early) (Fig. 6f, right) and in other donors/cultures (D1C1, D1C2, D1C3, D2C3, D3C2) (Fig. 3f). Moreover, these pathways were upregulated in winning subclones from independent barcoded donors/cultures (Extended Data Figs. 7–9), including the divergent subclone (ECB-1a, green) in D2C1R2 (Fig. 5c and Extended Data Fig. 8f), emphasizing their reproducibility. More generally, strong concordance between winning subclone and non-barcoded late subclones was observed across the top ten altered gene sets irrespective of mycoplasma levels, antibiotic treatment and other sources of biological and technical variation (Supplementary Fig. 21 and Methods). Similarly, winning subclones clustered with late cultures, which exhibited malignant transcriptional states based on unsupervised LSI projection (D1C1, D2C2, D2C3 and D3C2) (Fig. 6h and Extended Data Fig.

10a,b). Notably, there was a significant difference in the activation of p53, apoptosis and TNF signalling via NF-κB pathways (Fisher's exact test, Bonferroni corrected $P < 0.05$) between late (relative to early) and winning subclones compared with all other subclones (Extended Data Fig. 10c). These data highlight convergent phenotypic evolution in which the early activation of specific pathways is selectively advantageous, canalizing cells towards malignancy.

Discussion

Through multiyear experimental evolution of *TP53*-deficient HGO cultures, we model preneoplastic evolution and genotype–phenotype relationships following this common initiating insult. Remarkably, *TP53* deficiency was sufficient to recapitulate multiple hallmarks of CIN GC including aneuploidy, specific CNAs, SVs and transcriptional programmes, emphasizing the importance of cell-intrinsic processes during premalignant evolution. Although aneuploidy propagates heterogenous evolution, our data show preferred orders in the acquisition of CNAs, with early loss of chr3p and 9p frequently followed by biallelic inactivation of *CDKN2A* and/or *FHIT* and relatively late gain of 20q. Such preferred mutational orders have been described during tumorigenesis, most notably in the colon, but the resolution of inferences from cross-sectional data or established tumours is inherently limited[12,50]. Evolutionary phases in which deletions preceded whole-genome doubling and subsequent amplifications were recently reported in a murine model of *KrasG12D*, *Trp53*-deficient pancreatic cancer, but neither gene nor chromosome level orderings were seen in this system[51].

Our *TP53*$^{-/-}$ HGOs exhibited transcriptional and genomic hallmarks of premalignant gastro-oesophageal lesions despite remaining histologically normal. This is consistent with the requirement for genomic perturbation for even the earliest stages of gastro-oesophageal carcinogenesis and the accrual of complex rearrangements years before cancer diagnosis[15,16,23].

TP53$^{-/-}$ HGOs appear to be on a trajectory similar to *TP53*-deficient Barrett's oesophagus, for which the presumed cell of origin is gastric cardia[17], and proposed biomarkers of progression to oesophageal cancer include CNA acquisition and SV burden[16,52]. These in vitro models thus

recapitulate occult preneoplasia and mirror the latency of human tumorigenesis, with additional time or in vivo selective pressures evidently required for malignant transformation and further features of invasive disease such as whole-genome doubling or *ERBB2* amplification[22].

The finding that *TP53* deficiency elicits a temporally defined order of genomic aberrations raises the possibility that these features may similarly predict progression to CIN GC. Future evaluation of this hypothesis will require annotated intestinal metaplasia tissue collection with long-term follow-up. Although *TP53* deficiency elicits tissue-specific alterations that may aid in the detection of high-risk lesions, this constrained evolutionary state is unlikely to persist indefinitely given ensuing genome instability, emphasizing the need for earlier detection.

By joint measurement of lineage, CNAs and transcriptional states in individual cells, we investigated the molecular basis of clonal expansions and fitness. This showed stringent selection and reproducible subclonal dynamics across replicate cultures in which the same, initially rare, subclone fixed in the population. Pervasive clonal interference was evident amongst subclones, accompanied by intermittent periods of relative stasis, suggesting that an optimal karyotype has yet to be achieved, as reported in colorectal adenoma[53]. Furthermore, we observed a marked degree of phenotypic convergence on common dominant pathways across cultures and donors, irrespective of mycoplasma infection and antibiotic treatment. This evolutionary reproducibility is particularly notable given these and other potential sources of technical and biological variation, and implies that any such effects are evidently modest relative to the overwhelmingly dominant effect of *TP53* inactivation.

These first in-kind measurements address open questions concerning selection and determinism in clonal evolution extendable to other tissues. In the vast space of initiating insults, recurrent tissue-specific alterations can be prioritized to identify selectively advantageous alterations, temporal order constraints and convergent phenotypes. Such constraints, due to epistasis, can show barriers to malignant transformation and potential therapeutic targets. We anticipate that our results will advance empirical and theoretical investigations of mutation, selection and genome instability in human cells, much as the long-term evolution experiments pioneered by Lenski and

colleagues decades ago continue to yield fundamental insights into microbial adaptation[3,4].

Methods

A detailed description of the Materials and Methods is available in the Supplementary Information.

Reagent availability

Requests for reagents should be directed to the corresponding author.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Metadata and cellranger outputs are available at Zenodo (https://doi.org/10.5281/zenodo.6401895). Expressed cellular barcodes (ECB) sequencing data are available at bioProject ID (PRJNA838456). Genomic sequencing and scRNA-seq data are available at dbGAP under accession no. phs003249.v1. Source data are provided with this paper.

Code availability

The computational methods, procedures and analyses summarized above are implemented in custom R and python, and bash scripts are available via the Curtis Lab: https://github.com/cancersysbio/gastric_organoid_evolution.

References

1. Crosby, D. et al. Early detection of cancer. *Science* **375**, eaay9040 (2022).

2. Vázquez-García, I. et al. Clonal heterogeneity influences the fate of new adaptive mutations. *Cell Rep.* **21**, 732–744 (2017).

3. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).

4. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).

5. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).

6. Sun, R. et al. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* **49**, 1015–1024 (2017).

7. Laconi, E., Marongiu, F. & DeGregori, J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *Br. J. Cancer* **122**, 943–952 (2020).

8. Tsao, J. L. et al. Genetic reconstruction of individual colorectal tumor histories. *Proc. Natl Acad. Sci. USA* **97**, 1236–1241 (2000).

9. Baker, A.-M., Graham, T. A. & Wright, N. A. Pre-tumour clones, periodic selection and clonal interference in the origin and progression of gastrointestinal cancer: potential for biomarker development. *J. Pathol.* **229**, 502–514 (2013).

10. Williams, M. J. et al. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**, 895–903 (2018).

11. Rogers, Z. N. et al. Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat. Genet.* **50**, 483–486 (2018).

12. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

13. Waddingham, W. et al. Recent advances in the detection and management of early gastric cancer and its precursors. *Frontline Gastroenterol.* **12**, 322–331 (2021).

14. Mysuru Shivanna, L. & Urooj, A. A review on dietary and non-dietary risk factors associated with gastrointestinal cancer. *J. Gastrointest. Cancer* **47**, 247–254 (2016).

15. Li, X. et al. Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. *Cancer Prev. Res.* **7**, 114–127 (2014).

16. Paulson, T. G. et al. Somatic whole genome dynamics of precancer in Barrett's esophagus reveals features associated with disease progression. *Nat. Commun.* **13**, 2300 (2022).

17. Nowicki-Osuch, K. et al. Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science* **373**, 760–767 (2021).

18. Yan, H. H. N. et al. A comprehensive human gastric cancer organoid biobank captures tumor subtype heterogeneity and enables therapeutic screening. *Cell Stem Cell* **23**, 882–897 (2018).

19. Sethi, N. S. et al. Early TP53 alterations engage environmental exposures to promote gastric premalignancy in an integrative mouse model. *Nat. Genet.* **52**, 219–230 (2020).

20. Seidlitz, T., Koo, B.-K. & Stange, D. E. Gastric organoids-an in vitro model system for the study of gastric development and road to personalized medicine. *Cell Death Differ.* **28**, 68–83 (2021).

21. Lo, Y.-H. et al. A CRISPR/Cas9-engineered ARID1A-deficient human gastric cancer organoid model reveals essential and nonessential modes of oncogenic transformation. *Cancer Discov.* **11**, 1562–1581 (2021).

22. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).

23. Wang, K. et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582 (2014).

24. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).

25. Narkar, A. et al. On the role of p53 in the cellular response to aneuploidy. *Cell Rep.* **34**, 108892 (2021).

26. Weiss, M. B. et al. Deletion of p53 in human mammary epithelial cells causes chromosomal instability and altered therapeutic response. *Oncogene* **29**, 4715–4724 (2010).

27. Drost, J. et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).

28. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689 (2018).

29. Salehi, S. et al. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* **595**, 585–590 (2021).

30. Barrett, M. T. et al. Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat. Genet.* **22**, 106–109 (1999).

31. Saldivar, J. C. & Park, D. Mechanisms shaping the mutational landscape of the FRA3B/FHIT-deficient cancer genome. *Genes Chromosomes Cancer* **58**, 317–323 (2019).

32. Newell, F. et al. Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Med. Genomics* **12**, 31 (2019).

33. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

34. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–

210 (2020).

35. Glover, T. W., Wilson, T. E. & Arlt, M. F. Fragile sites in cancer: more than meets the eye. *Nat. Rev. Cancer* **17**, 489–501 (2017).

36. Birchenough, G. M. H., EV Johansson, M., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. *Mucosal Immunol.* **8**, 712–719 (2015).

37. Dong, D., Mu, Z., Zhao, C. & Sun, M. ZFAS1: a novel tumor-related long non-coding RNA. *Cancer Cell Int.* **18**, 125 (2018).

38. Rao, S. et al. β2-spectrin (SPTBN1) as a therapeutic target for diet-induced liver disease and preventing cancer development. *Sci. Transl Med.* **13**, eabk2267 (2021).

39. Paludan, S. R., Reinert, L. S. & Hornung, V. DNA-stimulated cell death: implications for host defence, inflammatory diseases and cancer. *Nat. Rev. Immunol.* **19**, 141–153 (2019).

40. Zhang, M. et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* **70**, 464–475 (2021).

41. Sathe, A. et al. Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. *Clin. Cancer Res.* **26**, 2640–2653 (2020).

42. Zhang, P. et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. *Cell Rep.* **30**, 4317 (2020).

43. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).

44. Levy, S. F. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).

45. Nguyen Ba, A. N. et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494–499 (2019).

46. Fujimoto, K. et al. Regulation of intestinal homeostasis by the ulcerative colitis-associated gene RNF186. *Mucosal Immunol.* **10**, 446–459 (2017).

47. Takeno, A. et al. Gene expression profile prospectively predicts peritoneal relapse after curative surgery of gastric cancer. *Ann. Surg. Oncol.* **17**, 1033–1042 (2010).

48. Maity, A. K. et al. Novel epigenetic network biomarkers for early detection of esophageal cancer. *Clin. Epigenetics* **14**, 23 (2022).

49. You, X. et al. Galectin-1 promotes vasculogenic mimicry in gastric adenocarcinoma via the Hedgehog/GLI signaling pathway. *Aging* **12**, 21837–21853 (2020).

50. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

51. Baslan, T. et al. Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).

52. Killcoyne, S. et al. Genomic copy number predicts esophageal cancer years before transformation. *Nat. Med.* **26**, 1726–1732 (2020).

53. Cross, W. et al. The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* **2**, 1661–1672 (2018).

Acknowledgements

We thank Z. Hu, S. Tilk, L. Attardi and A. Bhatt for helpful discussions, the Stanford University Hospital Tissue Procurement Shared Resource facility for specimen procurement and the Stanford Functional Genomics Core for assistance with sequencing. This work was supported by the US Department of Health & Human Services National Institutes of Health Director's Pioneer

Award (no. DP1-CA238296) to C.C. and a National Cancer Institute Cancer Target Discovery and Development Center (no. U01-CA217851) to C.J.K. and C.C. K. Karlsson was supported in part by a Swedish Research Council (Ventenskapsradet) International postdoctoral grant (no. 2018-00454).

Author information

Author notes

1. Moritz J. Przybilla

   Present address: Wellcome Sanger Institute & University of Cambridge, Hinxton, UK

2. These authors contributed equally: Kasper Karlsson, Moritz J. Przybilla

Authors and Affiliations

1. Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

   Kasper Karlsson, Hang Xu, Amanda Mah, Yuan-Hung Lo, Kathleen E. Houlahan, Calvin J. Kuo & Christina Curtis

2. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

   Kasper Karlsson, Eran Kotler, Hang Xu, Alexandra Sockell, Kathleen E. Houlahan & Christina Curtis

3. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

   Kasper Karlsson, Moritz J. Przybilla, Eran Kotler, Aziz Khan, Hang Xu, Kremena Karagyozova, Wing H. Wong, Katherine Liu, Kathleen E. Houlahan, Zhicheng Ma & Christina Curtis

4. Science for Life Laboratory and Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden

Kasper Karlsson

5. Department of Biology, Stanford University School of Medicine, Stanford, CA, USA

Katherine Liu

6. Department of Cell and Developmental Biology, University College London, London, UK

Bingxin Lu & Chris P. Barnes

7. Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Carlos J. Suarez

8. Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA

Christina Curtis

9. Chan Zuckerberg Biohub - San Francisco, San Francisco, CA, USA

Christina Curtis

Contributions

Conceptualization was the responsibility of C.C. Study design was carried out by K. Karlsson and C.C. K. Karlsson, K. Karagyozova, A.S. and A.M. performed organoid culturing. K. Karlsson, K. Karagyozova, A.S. and Z.M. carried out sequencing and library preparation. K. Karlsson, K. Karagyozova, W.H.W., A.M., Y.-H.L. and C.J.S. performed histology. Development of expressed cellular barcodes was undertaken by K. Karlsson and C.C. E.K. performed growth curve derivation estimation. Analysis of sWGS and WGS data was the responsibility of K. Karlsson, M.J.P., A.K.,

H.X., B.L., C.P.B. and C.C. Analysis of scRNA-seq data was carried out by K. Karlsson, M.J.P. and C.C. Analysis of ECB barcode data was undertaken by K. Karlsson, M.J.P., E.K., K.L. and C.C. K. Karlsson, M.J.P., H.X., E.K., A.K., K.E.H. and C.C. performed visualization. Funding acquisition was undertaken by C.J.K. and C.C. C.C. carried out project administration. C.J.K. and C.C. supervised the work. Writing of the original draft was done by K. Karlsson, M.J.P. and C.C. Writing review and editing were performed by all authors.

Corresponding author

Correspondence to Christina Curtis.

Ethics declarations

Competing interests

Unrelated to this study, C.J.K. is a founder and stockholder for Surrozen, Inc., Mozart Therapeutics and NextVivo, Inc. and C.C. is a stockholder in Illumina and DeepCell and an advisor to DeepCell, Genentech, Bristol Myers Squibb, 3T Biosciences, NanoString and ResistanceBio. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks Hans Clevers, James DeGregori, Toshiro Sato and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Schematic overview of gastric organoid cultures, assays and sequencing time points.

Organoids were established from three donors (abbreviated D) as wild-type (WT) cultures. For each donor (D1-D3), three independent CRISPR/Cas9 edited *TP53*$^{-/-}$ or *TP53*$^{-/-}$, *APC*$^{-/-}$ cultures (abbreviated C) were established (indicated by sg 1-3) and referred to as C1-C3 (Methods). The WT and genome edited cultures were evolved under defined conditions for over two years. Sequencing was performed across the experimental time course at defined intervals: *Early* (~0–200 days), *Mid* (~200–400 days), *Late* (~400–900 days). Each original culture was thawed (indicated by dashed lines) at an *Early/Mid* (190–290 days) and *Late* (540–730 days) time point for additional replication and comparisons. The thawed samples were treated with normocin to eliminate mycoplasma (Methods). All cultures were subject to shallow WGS (sWGS). A subset of cultures underwent deeper WGS and/or single cell RNA (scRNA)-sequencing at select time points. In addition to these non-barcoded cultures, representative *TP53*$^{-/-}$ HGO cultures from each donor were selected for prospective lineage tracing via transduction of a lentiviral expressed cellular barcode (ECB), as indicated by the multi-coloured circle in the legend. These ECB cultures were similarly subject to sWGS and scRNA-seq. Broad time intervals are indicated as in the legend, while days in culture are provided for individual cultures. Note that scRNA for D1C3 “Mid trajectory” was sampled at day 413.

Extended Data Fig. 2 Recurrent copy number aberrations in *TP53*-deficient gastric organoids are enriched in gastric and esophageal cancers.

a, Prevalence of somatic alterations and fraction genome altered in gastric cancer (stomach adenocarcinoma, STAD) from TCGA in all subgroups versus the CIN subgroup. Data derive from the cBioPortal. **b**, Frequency of chromosome arm alterations in *TP53*$^{-/-}$ HGOs (ORG) at late time points (days 588 to 835, as in Fig. 1c) relative to other tumour types: Stomach Adenocarcinoma (STAD), Esophageal carcinoma (ESCA), Colorectal

Adenocarcinoma (COAD), Rectum Adenocarcinoma (READ), Breast invasive carcinoma (BRCA), Glioblastoma Multiforme (GBM), Pancreatic Adenocarcinoma (PAAD). TCGA data were obtained from Firehose (http://gdac.broadinstitute.org/#). **c**, Enrichment of chromosome arm-level alterations in *TP53*$^{-/-}$ gastric organoid cultures across cancer types. Boxes show inter-quartile range (IQR), center lines represent the median, whiskers extend by 1.5 × IQR. Arm-level CNAs altered in two or more *TP53*$^{-/-}$ gastric organoid cultures (n = 12 alterations) were more frequently altered than alterations present in 1 or fewer cultures (n = 66 alterations) in both STAD and ESCA (p-value shown, two-sided Wilcoxon rank sum test). **d**, The Jaccard Index (JI) was calculated by comparing CNAs that occurred in more than 1 chromosome arm in organoid cultures with CNAs occurring in 15% or more cases in a given tumour type. Permutation tests were performed to determine if the JI score was higher than expected by chance (i.e. one-sided test). For each tumour type, organoid CNA labels were randomly sampled (n = 10,000) from all possible chromosome arm-level events, and the JI was calculated. The empirical p-value was calculated as: P = 1-(sum(real JI > null JIs)/number of null JIs). Boxplot represents median, 0.25 and 0.75 quantiles with whiskers at 1.5 × IQR and includes the nominal p-values. **e**, Oncoplot shows alterations that occurred in two or more samples for genes commonly (>10% of cases) altered in gastro-esophageal cancer.

Source Data

Extended Data Fig. 3 Longitudinal whole genome sequencing (WGS) of *TP53*$^{-/-}$ gastric organoids .

a, Overview of WGS and scRNA-seq time points for Early, Mid and Late cultures. Time is indicated in days (d). **b**, Summary of genomic features as assessed by WGS of multiple time points for Donors 2 and 3, expanding upon Fig. 2a. *Mid* and *Late* time points correspond to day 296 and 705–754, respectively. **c**, Distribution of non-clustered SVs, simple SVs (2–9 rearrangements) and complex SVs (10 or more rearrangements) defined using ClusterSV (Methods). **d**, Distribution of SV types across the three classes of SVs (non-clustered, clustered-simple and clustered-complex) **e**, Boxplot comparing total SV burden at the time of endoscopy (T1 and T2)

for four Barrett's esophagus biopsies per patient with cancer outcome (CO, n = 160) or noncancer outcome (NCO; n = 160) from the Paulson et al. cohort relative to the total SV burden between early (n = 4) and late (n = 9) timepoints in *TP53*$^{-/-}$ and *TP53*$^{-/-}$, *APC*$^{-/-}$ HGOs (right). P-values were calculated using Wilcoxon rank sum test (two sided, unpaired). Boxes show IQR, center lines represent the median, whiskers extend by 1.5 × IQR.

Extended Data Fig. 4 *TP53*-deficient gastric organoids recapitulate complex structural variants (SVs) observed in gastric cancers.

a, Complex rigma-like SVs seen in gastric cancer (GC) patients such as pfg008 from the Wang et al. cohort (Methods) are similar to those in *TP53*$^{-/-}$ HGOs including D3C1 (also shown in Fig. 2e). **b**, Zoomed-in view of a region on chromosome 3 which evolved complex SVs during *in vitro* culture. **c**, Corresponding CNA profiles based on longitudinal WGS of D2C2 at 5 timepoints spanning days 115 to 729 in culture. **d**, Fishplot for D2C2 depicting CNA evolution inferred from WGS. **e**, IGV plots indicate the translocation between chromosomes 3 and 9 for D2C2, corresponding to the complex SV in panel b. Primers used for PCR amplification, gel electrophoresis purification and Sanger sequencing as marked in the plots **f**, BLAT results from Sanger sequencing of the PCR primers demonstrating partial alignment to chr3 and chr9. Image of stomach in **d** is from Servier Medical Art, CC BY 3.0.

Source Data

Extended Data Fig. 5 Latent Semantic Index (LSI) projection of gastric organoids onto gastric tissue dataset.

a–b, UMAP visualizations coloured according to culture (left) and time point (right) for Donors 2 and 3 depicting 9,031 and 8,591 cells, respectively. **c–d**, Dotplot depicting the expression of selected marker genes for individual cultures and time points. Coloured bars highlight marker genes associated with normal gastric and intestinal cell types, genes up-regulated in the gene expression profiling interactive analysis (GEPIA) of GC, and others of functional relevance. Pit mucosal cells: *MUC5AC*, *TFF1* –

dark yellow; Gland mucosal cells: *MUC6*, *TFF2* – light blue; Proliferative cells: *MKI67* – purple; Neck-like cells: *PGC*, *LYZ* – orange; Mucosal stem cells: *OLFM4* – turquoise; Enterocytes: *FABP1*, *VIL1* – olive; Goblet cells: *TFF3*, *WFDC2*, *MUC5B*, *CDX2* – green; GEPIA top 12 genes: *CEACAM5*, *CEACAM6*, *CLDN3*, *CLDN4*, *CLDN7*, *REG4*, *MUC3A*, *MUC13*, *PI3*, *UBD*, *AOC1*, *CDH17* – black; Other: *TP53*, *APC*, *CDKN2A*, *FHIT* – red. **e**, The reference gastric tumour-normal dataset (Sathe et al.) for the LSI projection is shown with key cellular populations coloured according to their expression phenotype, as denoted in the legend. This panel is identical to that shown in Fig. 4c. **f–h**, Showing the LSI projection of individual cultures for donors 1, 2 and 3.

Extended Data Fig. 6 Subclone growth and frequency comparison and barcode trajectories for D1C1 and D2C3.

a, sWGS of the parental population (top) and three time points for D1C1 and D2C3, replicates 1-3. **b**, Muller plots of the ECB subclone frequency over time for the cultures shown in panel a reveals similar lineage dynamics across replicates and deterministic outgrowth. **c**, Barplot depicting frequencies of the top 10 subclones in the parental population and the winning subclone (which was not one of the top subclones in the parental population). All other subclones coloured gray. **d**, Pairwise Pearson correlation over time between replicate cultures for D1C1, D2C1, D2C2, D2C3 and D3C2. Time points correspond to passage times, where intervals between passages are approximately two weeks.

Source Data

Extended Data Fig. 7 Linking single-cell genotypes to their transcriptional phenotypes in D2C1R1.

a, Inferred copy number heatmap from scRNA-seq data at day 245, analogous to Fig. 6a. **b**, Copy number plot of parent population for D2C1. **c**, Fishplot schematic illustrating the link between lineage (ECBs) and copy number based subclones, similar to Fig. 6b. The actual subclone frequencies are shown in Fig. 5c. **d**, Scatterplot comparing subclone frequencies at day 157 and the fold change between days 129 and 157 for all subclones.

Subclones of interest are highlighted as in panel a. **e**, Dotplot showing the expression of top differentially expressed genes (DEGs) based on GEPIA (gene expression profiling interactive analysis) of gastric cancer (GC). **f**, Volcano plot illustrating DEGs from the comparison of the winning subclone 0a and its parent subclone 0b. Vertical and horizontal lines correspond to absolute log2FC values of 1 and p-values < 0.00001 (Wilcoxon rank sum test, not corrected for multiple testing) respectively. **g**, GSEA heatmap from MSigDB Hallmark gene sets showing the most significantly altered pathways for the highlighted subclones (right; Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted, two-sided). A manually reconstructed copy number phylogeny is shown above. This visualization is equivalent to Fig. 6f.

Source Data

Extended Data Fig. 8 Linking single-cell genotypes to their transcriptional phenotypes in in D2C1R2.

a, Inferred copy number heatmap from scRNA-seq data at day (d) 245, analogous to Fig. 6a. **b**, Fishplot schematic illustrating the link between lineage (ECBs) and copy number based subclones, similar to Fig. 6b. The actual subclone frequencies are shown in Fig. 5c. **c**, Scatterplot comparing subclone frequencies at day 157 and the fold change between days 129 and 157 for all subclones. Subclones of interest are highlighted as in panel a. **d**, Dotplot showing the expression of top differentially expressed genes (DEGs) based on GEPIA (gene expression profiling interactive analysis) of gastric cancer (GC). **e**, Volcano plot illustrating DEGs from the comparison of the winning subclone 0a and its parent subclone 0b. Vertical and horizontal lines correspond to absolute log2FC values of 1 and p-values < 0.00001 (Wilcoxon rank sum test, not corrected for multiple testing) respectively. **f**, GSEA heatmap from MSigDB Hallmark gene sets showing the most significantly altered pathways for the highlighted subclones (right; Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted, two-sided). A manually reconstructed copy number phylogeny is shown above. This visualization is equivalent to Fig. 6f.

Source Data

Extended Data Fig. 9 Linking single-cell genotypes to their transcriptional phenotypes in in D3C2R1.

a, Inferred copy number heatmap from scRNA-seq data at day (d) 189, analogous to Fig. 6a. **b**, Fishplot schematic illustrating the link between lineage (ECBs) and copy number based subclones, similar to Fig. 6b. The actual subclone frequencies are shown in Fig. 5c. **c**, Scatterplot comparing subclone frequencies at day 157 and the fold change between days 129 and 157 for all subclones. Subclones of interest are highlighted as in panel A. **d**, Dotplot showing the expression of top differentially expressed genes (DEGs) based on GEPIA (gene expression profiling interactive analysis) of gastric cancer (GC). **e**, Volcano plot illustrating DEGs from the comparison of the winning subclone 0a and its parent subclone 0b. Vertical and horizontal lines correspond to absolute log2FC values of 1 and p-values < 0.00001 (Wilcoxon rank sum test, not corrected for multiple testing) respectively. **f**, GSEA heatmap from MSigDB Hallmark gene sets showing the most significantly altered pathways for the highlighted subclones (right; Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted, two-sided). A manually reconstructed copy number phylogeny is shown above. The visualization is equivalent to Fig. 6f.

Source Data

Extended Data Fig. 10 Enrichment of similar gene sets between winning subclones and *Late* cultures.

a, GSEA heatmap from MSigDB Hallmark gene sets for each barcoded culture, including replicates: D2C2-R2 (day 173), D2C1R1 (day 245), D2C1-R2 (day 245), D3C2R1 (day 189), as well as non-barcoded *Late* cultures for D1C1, D1C2, D1C3, D2C2, D2C3 and D3C2. The GSEA score is indicated by dot size and coloured according to the directionality of expression profiles (up, red; down, blue). Background shading indicates statistical significance (Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted, two-sided) **b**, Winning subclones and *Late* samples cluster together based on GSEA score. Pairwise spearman correlation between samples based on GSEA scores for the top 10 most altered pathways for all *Late* and winnings subclone samples. The most altered pathways are shown in a. This

plot is identical to Fig. 6g but with sample annotations included. Note that the winning subclones (marked in red) cluster with the *Late* samples for D1C1, D2C2, D2C3 and D3C2 which exhibit a more malignant phenotype compared to *Late* samples D1C2 and D1C3 based on the LSI projection (Fig. 4c,d, Extended Data Fig. 5). **c**, Fisher exact test of independence (two-sided), comparing significance for each pathway, and status for each sample - *Late* and winning subclones (n = 16) relative to all other subclones analysed (n = 29). The red line indicates the significance threshold (0.05) with Bonferroni correction.

Source Data

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–21.

Reporting Summary

Supplementary Table 1

List of oligonucleotides utilized in this study.

Supplementary Table 2

Cell passaging information for all donors and cultures.

Supplementary Table 3

Summary of sWGS data including time points, coverage and quality control metrics and genomic alterations.

Supplementary Table 4

Summary of WGS data including time points, coverage and quality control and ploidy estimates.

Supplementary Table 5

CNAs accompanying fishplot schematics.

Supplementary Table 6

scRNA sequencing results including quality control metrics, top differentially expressed genes and GSEA results.

Supplementary Table 7

Timing of clonal sweeps based on ECB sequencing data.

Supplementary Table 8

Subclone frequencies in fishplot schematics.

Supplementary Table 9

scRNA sequencing information for ECB subclones, top differentially expressed genes and GSEA results.

Source data

Source Data Figs. 1–3 and 5–6.

Source Data Extended Data Figs. 2, 4 and 6–10.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation,

distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Karlsson, K., Przybilla, M.J., Kotler, E. *et al.* Deterministic evolution and stringent selection during preneoplasia. *Nature* **618**, 383–393 (2023). https://doi.org/10.1038/s41586-023-06102-8

- Received: 01 September 2022
- Accepted: 19 April 2023
- Published: 31 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06102-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

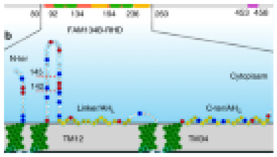
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06102-8

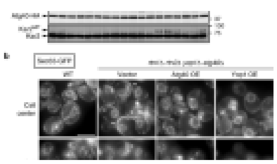
| Section menu | Main menu |



Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy

30 May 2019

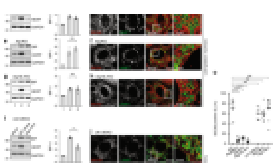
Ramachandra M. Bhaskara, Paolo Grumati, … Gerhard Hummer



Super-assembly of ER-phagy receptor Atg40 induces local ER remodeling at contacts with forming autophagosomal membranes

03 July 2020

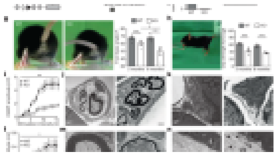
Keisuke Mochida, Akinori Yamasaki, … Hitoshi Nakatogawa



ESCRT-III-driven piecemeal micro-ER-phagy remodels the ER during recovery from ER stress

07 November 2019

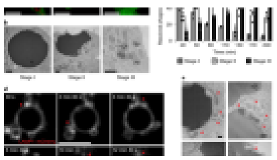
Marisa Loi, Andrea Raimondi, … Maurizio Molinari



[Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy]

24 May 2023

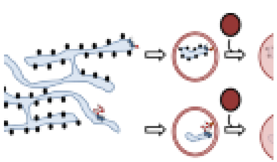
Hector Foronda, Yangxue Fu, … Christian A. Hübner



[Phagolysosome resolution requires contacts with the endoplasmic reticulum and phosphatidylinositol-4-phosphate signalling]

30 September 2019

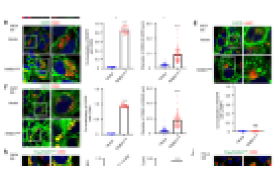
Roni Levin-Konigsberg, Fernando Montaño-Rendón, … Sergio Grinstein



[ER-phagy and human diseases]

28 October 2019

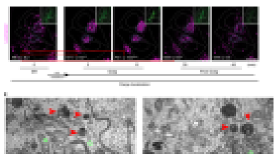
Christian A. Hübner & Ivan Dikic



RAB31 marks and controls an ESCRT-independent exosome pathway

21 September 2020

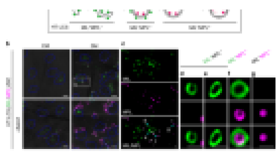
Denghui Wei, Weixiang Zhan, … Tiebang Kang



KDEL receptor regulates secretion by lysosome relocation- and autophagy-dependent modulation of lipid-droplet turnover

13 February 2019

Diego Tapia, Tomás Jiménez, … Jorge Cancino



An autophagy assay reveals the ESCRT-III component CHMP2A as a regulator of phagophore closure

20 July 2018

Yoshinori Takahashi, Haiyan He, … Hong-Gang Wang

- Article
- Open Access
- Published: 24 May 2023

Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum

- Alexis González,
- Adriana Covarrubias-Pinto,
- Ramachandra M. Bhaskara,
- Marius Glogger,
- Santosh K. Kuncha,
- Audrey Xavier,
- Eric Seemann,
- Mohit Misra,
- Marina E. Hoffmann,
- Bastian Bräuning,
- Ashwin Balakrishnan,
- Britta Qualmann,
- Volker Dötsch,
- Brenda A. Schulman,
- Michael M. Kessels,
- Christian A. Hübner,
- Mike Heilemann,
- Gerhard Hummer &
- …
- Ivan Dikić

Show authors

Nature **volume 618**, pages 394–401 (2023)

- 12k Accesses

- 218 Altmetric

- Metrics details

Abstract

The endoplasmic reticulum (ER) undergoes continuous remodelling via a selective autophagy pathway, known as ER-phagy[1]. ER-phagy receptors have a central role in this process[2], but the regulatory mechanism remains largely unknown. Here we report that ubiquitination of the ER-phagy receptor FAM134B within its reticulon homology domain (RHD) promotes receptor clustering and binding to lipidated LC3B, thereby stimulating ER-phagy. Molecular dynamics (MD) simulations showed how ubiquitination perturbs the RHD structure in model bilayers and enhances membrane curvature induction. Ubiquitin molecules on RHDs mediate interactions between neighbouring RHDs to form dense receptor clusters that facilitate the large-scale remodelling of lipid bilayers. Membrane remodelling was reconstituted in vitro with liposomes and ubiquitinated FAM134B. Using super-resolution microscopy, we discovered FAM134B nanoclusters and microclusters in cells. Quantitative image analysis revealed a ubiquitin-mediated increase in FAM134B oligomerization and cluster size. We found that the E3 ligase AMFR, within multimeric ER-phagy receptor clusters, catalyses FAM134B ubiquitination and regulates the dynamic flux of ER-phagy. Our results show that ubiquitination enhances RHD functions via receptor clustering, facilitates ER-phagy and controls ER remodelling in response to cellular demands.

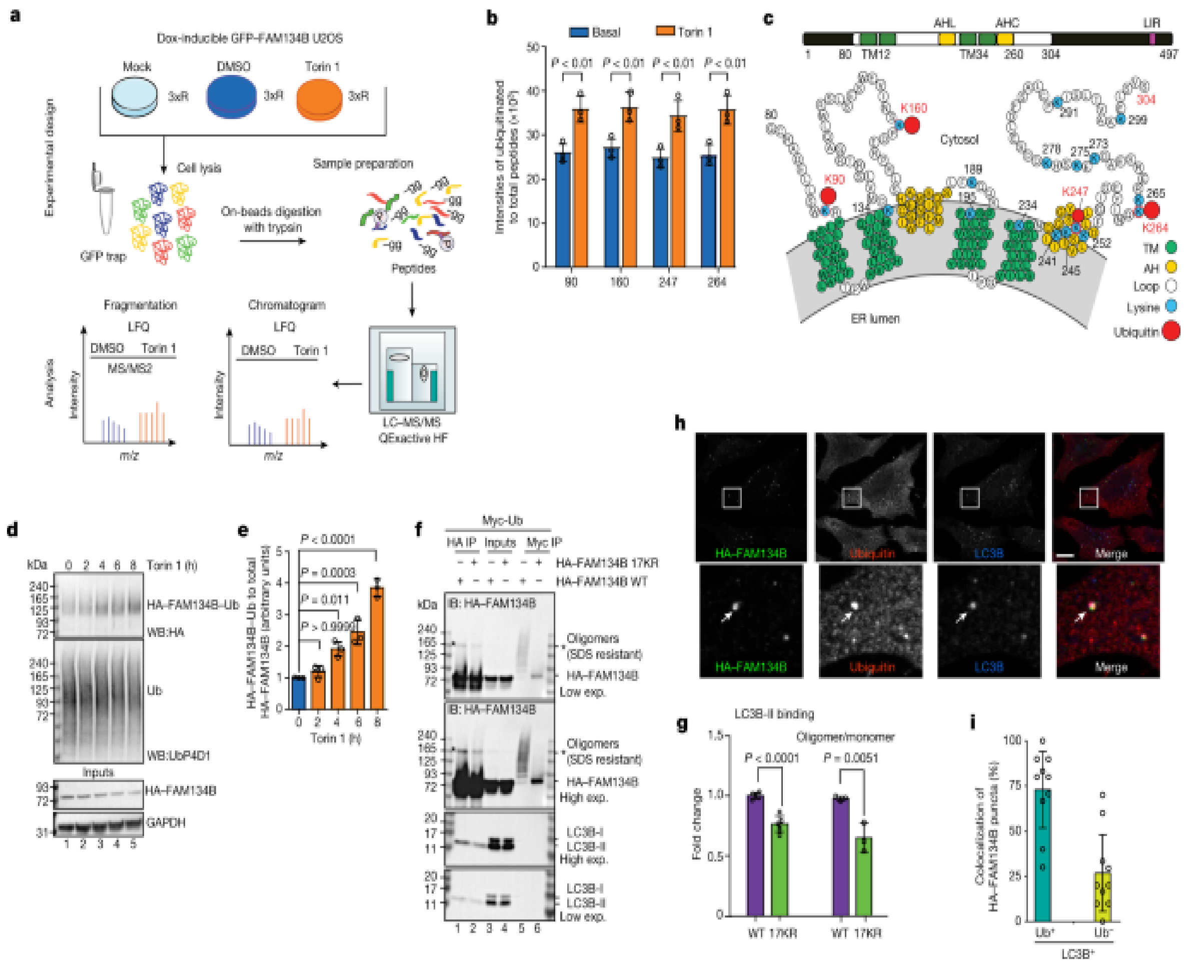
Main

FAM134B is a mammalian reticulon-like protein that shapes the ER membrane[3,4]. It is also an ER-phagy receptor, mediating the fragmentation and selective degradation of ER sheets[3]. Structural modelling and molecular simulations have revealed that the RHD of FAM134B forms wedge-shaped membrane inclusions that induce positive membrane curvature to promote ER fragmentation, assisted by RHD clustering[4,5,6]. However, it is unclear how FAM134B-mediated ER-phagy is regulated in mammalian cells.

FAM134B is ubiquitinated at the RHD

Ubiquitination regulates a large number of cellular processes, so we used mass spectrometry (MS) to investigate its potential role in ER-phagy by mapping the ubiquitination profile of FAM134B (Fig. 1a). Proteomic analysis of FAM134B-derived diGly peptides identified residues K90, K160, K264 and K247 as the primary ubiquitination sites (Fig. 1b), which are located within the cytosolic segments of the FAM134B RHD (Fig. 1c). The induction of ER-phagy with Torin 1 increased the representation of all four diGly peptides (Fig. 1b) as well as the overall endogenous ubiquitination level of haemagglutinin (HA)-tagged FAM134B (Fig. 1d,e and Extended Data Fig. 1a, compare lanes 3 and 1, and Extended Data Fig. 1b). The accumulation of total and ubiquitinated endogenous FAM134B in cells treated with bafilomycin A1 (BafA1) indicated lysosomal degradation (Extended Data Fig. 1a, compare lane 4 to lanes 3 and 2, and Extended Data Fig. 1b). No accumulation was observed in cells treated with the proteasome inhibitor MG132 (Extended Data Fig. 1c). Accordingly, cycloheximide chase experiments showed that BafA1 rendered FAM134B more stable (Extended Data Fig. 1d,e). We also found that TAK243, a potent inhibitor of the ubiquitin (Ub)-activating enzyme, abolished endogenous FAM134B ubiquitination (Extended Data Fig. 1f) and delayed its basal turnover (Extended Data Fig. 1g,h). As an alternative method, we substituted lysine residues identified by MS (K90, K160, K247 and K264) and their neighbours (K252, K265, K278 and K291) with arginine, but overall ubiquitination levels, binding to LC3B-II and the number of FAM134B–LC3B-decorated ER fragments were not affected in this mutant (Extended Data Fig. 1i,j). Only the replacement of the entire set of 17 highly conserved lysine residues (mutant HA–FAM134B17KR) resulted in a strong decrease in FAM134B RHD ubiquitination (Fig. 1f, compare lanes 6 and 5). Of note, the lack of FAM134B RHD ubiquitination correlated with decreased binding to LC3B-II (Fig. 1f, compare lanes 2 and 1, and Fig. 1g; approximately 1.32-fold reduction). High-molecular-weight (oligomeric) species of FAM134B were less abundant in the HA–FAM134B17KR mutant (Fig. 1f, compare lanes 2 and 1, and Fig. 1g; approximately 1.5-fold reduction). Similar results were observed following the chemical crosslinking of intact membranes from cells expressing the wild-type (HA–FAM134BWT) or mutant (HA–FAM134B17KR) receptor are consistent with the high-molecular-weight species representing oligomers (Extended Data Fig. 1k).

Fig. 1: Ubiquitination of FAM134B RHD.



a, MS workflow for the analysis of GFP–FAM134B co-immunoprecipitated from cell lysates following treatment with Torin 1 (6 h, orange) and DMSO (basal, blue) or from mock-treated cells (GFP empty, light blue). HF, Qexactive HF (Ultra-High-Field Orbitrap) mass spectrometer; LFQ, label free quantification; 3xR 3 biological replicates. Panel **a** was partly generated using Servier Medical Art (Servier), licensed under a Creative Commons Attribution 3.0 unported license. **b**, Ubiquitination of FAM134B under control conditions and in response to 250 nM Torin 1 for 6 h. The diGly peptide intensities of FAM134B are normalized to the total intensities of modified and non-modified FAM134B peptides (data are mean ± s.d.; $n = 3$ independent experiments, two-way ANOVA, Bonferroni post-hoc test). **c**, Schematic organization of FAM134B. The RHD consists of two transmembrane segments (TM, green) separated by a linker and two conserved amphipathic helices (AH, yellow). The conserved lysine residues (blue) and ubiquitinated lysines (red) are highlighted. **d**, TUBE-2 pulldown assay showing increased time-dependent ubiquitination of FAM134B

following Torin 1 treatment. WB, western blot. **e**, Densitometric quantification of the immunoblot signals for ubiquitinated HA–FAM134B normalized to total HA–FAM134B levels (**d**). Data are mean ± s.d.; $n = 3$ independent experiments; one-way ANOVA, Bonferroni post-hoc test. **f**, FAM134B RHD ubiquitination assay in cells showing a reduction of ubiquitination when the 17 conserved lysine residues are replaced with arginine (Myc–Ub immunoprecipitation (IP)). Lack of RHD ubiquitination reduces binding to LC3B-II and the abundance of oligomeric species (HA–FAM134B IP). IB, immunoblot; exp., experiment. **g**, Densitometric quantification of the immunoblot signals from **f** (lanes 1 and 2): LC3B-II bound to HA–FAM134B WT or HA–FAM134B 17KR, and the oligomers. Data are mean ± s.d.; $n = 5$ and $n = 3$ independent experiments, respectively; one-tailed unpaired Student's *t*-test. **h**, Confocal fluorescence microscopy analysis of HA–FAM134B co-labelled with LC3B and Ub in cells treated with 250 nM Torin 1 for 2 h. Arrows indicate autophagosome (LC3B positive in blue) that colocalizes with HA FAM134B (green) and ubiquitin (red). Scale bar, 10 µm. **i**, Quantification of the fluorescence signal of HA–FAM134B–LC3B puncta that colocalizes or not with Ub from **h** (based on Pearson's correlation coefficients). Data are mean ± s.d., $n = 10$ cells.

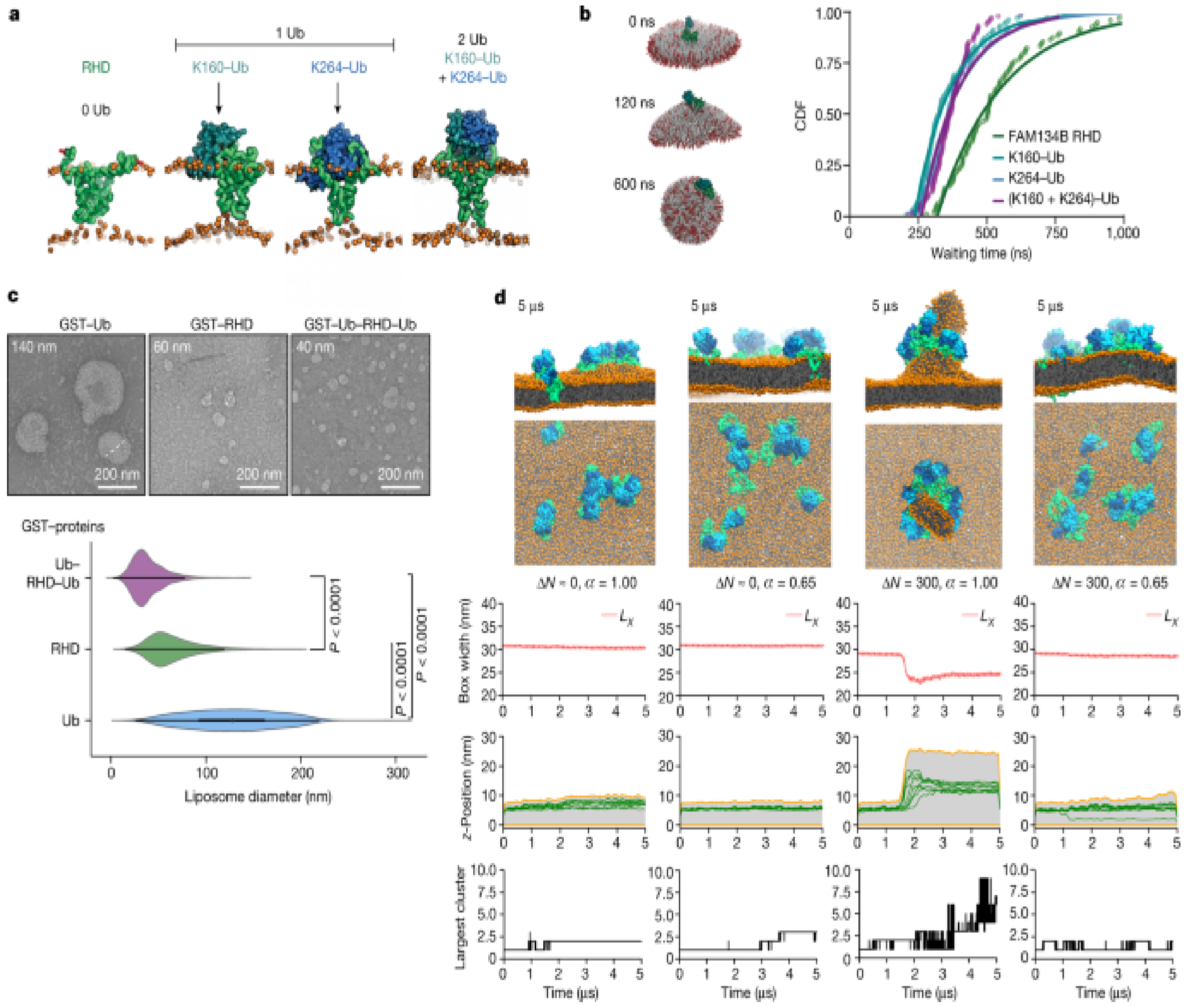
Source data

Several results indicated that ubiquitination of the FAM134 RHD can promote the formation and/or stabilization of FAM134B oligomers implicated in ER-phagy. First, high-molecular-weight species of FAM134B accumulate when the flux of ER-phagy is inhibited and/or following the deletion of the LC3-interacting region (LIR), indicating that FAM134B is delivered to lysosomes in its oligomeric form (Extended Data Fig. 1l,m). Second, in the absence of the FAM134B LIR, the ubiquitinated forms of FAM134B accumulate, confirming its autophagy-mediated degradation (Extended Data Fig. 1n). Third, immunofluorescence analysis following exposure to Torin 1 showed that approximately 70% of FAM134B$^+$ autophagosomes contained Ub, suggesting Ub has a wide-engaging role in FAM134B-driven ER-phagy (Fig. 1h,i). These results show that the ubiquitination of the FAM134 RHD can promote the formation and/or stabilization of FAM134B oligomers implicated in ER-phagy.

Ubiquitination enhances and fine tunes RHD membrane remodelling functions

We studied the effect of ubiquitination on RHD structure and dynamics with coarse-grained MD simulations of mono-ubiquitinated and bi-mono-ubiquitinated FAM134B RHD molecules, namely, K160–Ub, K264–Ub and (K160 + K264)–Ub, embedded in POPC (16:0/18:1 PC) lipid bilayers. The most populated conformations of the ubiquitinated FAM134B RHD molecules are shown in runs of up to 10 μs each (Fig. 2a). The Ub moieties perturb the RHD structure locally by mediating multiple interactions with the proximal cytosolic loops of the RHD and the POPC bilayer (Extended Data Fig. 2a–d). The main hydrophobic face of K160–Ub makes substantial contact with POPC lipids, widening the wedge shape of the RHD (Extended Data Fig. 2b). In the RHD of K264–Ub, Ub interacts primarily with the cytosolic loops and is located on top of the RHD (Extended Data Fig. 2c,d). In the bi-mono-ubiquitinated variant (K160 + K264)–Ub, the two Ub moieties mediate *cis*-interactions with each other and are bundled on top of the RHD (Extended Data Fig. 2e,f). Increased radii of gyration of ubiquitinated RHDs result in a larger footprint on the bilayer, thus possibly perturbing it more severely (Extended Data Fig. 2g,h). These changes in intrinsic RHD structure due to ubiquitination may affect its membrane curvature induction and sensing functions.

Fig. 2: Role of Ub in the structure and function of RHD.



a, Equilibrated structures of ubiquitinated and non-ubiquitinated FAM134B RHD variants in MD simulations highlight the arrangement of Ub moieties (cyan and blue) with respect to the RHD (green) and the POPC bilayer (orange beads). **b**, Ubiquitination accelerates membrane curvature induction. K160–Ub (left) induces faster transitions of bicelles to closed vesicles than non-ubiquitinated RHDs (cumulative distribution function (CDF) of waiting times shown on the right; compare cyan, blue and purple versus green). **c**, In vitro liposome remodelling experiments using purified protein samples (GST–Ub, GST–RHD$_{90–264}$, GST–Ub–RHD$_{90–264}$–Ub) incubated with liposomes for 8 h at 25 °C. The top panel shows representative negative-stain transmission electron micrographs. Remodelled proteoliposomes were quantified by measuring their diameters (dotted lines) using ImageJ (version 1.51w). The bottom panel shows violin plots of liposome size distributions. Violin plots show the boxplots with median value (black dot), the interquartile range (black shaded region), the minimum and maximum

values (1.5× the interquartile region) and mirrored probability density estimates on sides (coloured shaded region). GST–RHD mean = 64.40 nm; GST–Ub–$RHD_{90–264}$–Ub mean = 39.10 nm; GST–Ub mean = 127.47 nm. GST–$RHD_{90–264}$ $n = 625$; GST–Ub–$RHD_{90–264}$–Ub $n = 961$; GST–Ub $n =$ 1,573; Kruskal–Wallis or Dunn's post-hoc test. **d**, Ub promotes receptor clustering and membrane remodelling. Snapshots show the arrangement of nine (K160 + K264)–Ub–RHD molecules in model POPC bilayers (orange) at the end of the simulations. MD simulations were performed under four different conditions by altering bilayer asymmetry, $\Delta N = 0$ and 300, and protein–protein interaction strength, $\alpha = 1$ and 0.65, as quantified in the panels below. For the rows: time traces of the box width (L_X) during the four simulations (top); vertical displacement (z) of individual ubiquitinated proteins (centre-of-mass positions shown as green lines), with the highest and lowest points in the membrane shown as orange lines, and the intervening range in light grey (middle); the size of the largest protein cluster as a function of time for different simulation conditions (bottom).

Source data

To test whether ubiquitination affects the RHD-mediated induction of membrane curvature, we used in silico simulations to remodel a discontinuous bicelle (DMPC + DHPC lipids) into a closed vesicle[4] at 300 K (Fig. 2b and Supplementary Videos 1–3). The ubiquitinated and non-ubiquitinated RHDs induced a positive mean curvature (+H) leading to vesicle formation within the first 500 ns (3 × 100 replicates; 1 µs each; Extended Data Fig. 3a–d). By measuring the kinetics and rates for vesicle formation (see Methods; Extended Data Fig. 3e,f), we estimated that single K160–Ub, K264–Ub and (K160 + K264)–Ub species accelerated vesicle formation moderately by factors of 1.49, 1.37 and 1.38, respectively (Fig. 2b and Extended Data Fig. 3g). To highlight the effect of Ub on curvature induction, we performed these assays at 280 K, which increases the energy barrier to induce curvature and vesiculation. We found that bicelles containing the FAM134B RHD and K160–Ub did not effectively transit to closed vesicles (Extended Data Fig. 3h,k), whereas bicelles containing K264–Ub or (K160 + K264)–Ub formed closed vesicles, indicating that these variants can overcome the barrier even at 280 K (Extended Data Fig. 3i,j). Lateral diffusion of ubiquitinated variants in buckled bilayers showed

no differences in preferred curvature, indicating that the strong curvature-sensing function of the RHD is preserved upon ubiquitination (Extended Data Fig. 4a–c).

Next, we tested how ubiquitination affects the FAM134B RHD-mediated remodelling of liposomes in vitro. We created and purified N-terminal and C-terminal gene fusions of Ub to FAM134B $RHD_{90–264}$ (described as Ub–RHD–Ub). As a control, we created a construct encompassing the entire RHD of FAM134B without Ub ($RHD_{90–264}$). The incubation of liposomes with Ub–RHD–Ub led to significantly smaller structures with a narrow distribution than the liposomes incubated with RHD proteins alone (Fig. 2c), indicating a significant gain of membrane-remodelling activity for ubiquitinated RHD proteins. In cells, both constructs localized to the ER, based on the ER marker REEP5. We observed a significant increase in the number of RHD–REEP5-containing puncta in cells expressing Ub–RHD–Ub. These puncta may represent clusters of RHD-containing proteins, enhanced by the presence of ubiquitinated FAM134B RHDs (Extended Data Fig. 4d,e).

Ub interactions facilitate large-scale membrane deformations

Next, we simulated the behaviour of ubiquitinated RHDs to induce curvature-mediated protein sorting and clustering (Extended Data Fig. 4f). In the absence of ubiquitination, the two RHD molecules on the buckled membrane diffused to the top of the buckle, where they formed a loose cluster. Ubiquitination slowed down the diffusion somewhat but resulted in a tighter cluster at the top of the buckle that persisted for the entire simulation (up to 25 µs). From 19 µs onwards, Ub-mediated contacts stabilized the dimeric complex (inset in Extended Data Fig. 4f and Supplementary Video 4). Next, we simulated ten ubiquitinated FAM134B RHD molecules (five K160–Ub and five K264–Ub) embedded in a closed tubule (up to approximately 8.5 µs). The Ub moieties initiated *trans*-interactions among the RHDs and enabled the formation of three protein clusters (dimers and trimers) on the MD timescale (Supplementary Video 5). These clusters

deformed the tubule in both principal directions (squares in Extended Data Fig. 4g).

Motivated by these locally bud-shaped structures, we studied the effect of ubiquitination on spontaneous membrane budding by simulating nine bi-mono-ubiquitinated RHDs ((K160 + K264)-Ub) embedded in POPC bilayers under four different simulation conditions (Fig. 2d). We increased the lipid-number asymmetry of the bilayer leaflets from $\Delta N = 0$ to 300 to increase the energetic driving force for budding, and we reduced the protein–protein interaction (PPI) strength from 100% ($\alpha = 1$) to 65% ($\alpha = 0.65$)[6] to weaken Ub-mediated protein clustering. Spontaneous budding was observed in the asymmetric bilayer with 100% PPI strength (Fig. 2d and Supplementary Video 6), triggered by the formation of a cluster of ubiquitinated RHDs on top of the nascent bud. After budding, the remaining ubiquitinated RHDs sorted onto the membrane bud to form a Ub-rich protein coat in the form of a ring around its neck. Ub–Ub interactions are characteristic of these RHD clusters (blue and cyan moieties in Fig. 2d). By contrast, in all the other three simulations, RHD clusters only formed transiently (because of reduced PPI strength) and induced at most local membrane bulges (because of the high-energy penalty for budding at low asymmetry, $\Delta N = 0$). We conclude that the stabilization of RHD clusters by Ub-mediated interactions facilitates FAM134B-induced membrane budding.

Analysis of all possible Ub–Ub interactions (Extended Data Fig. 5a–d) indicated that intermolecular or *trans*-Ub–Ub interactions triggered protein clustering and oligomerization (Extended Data Fig. 5c). Although changing the membrane asymmetry and PPI strength during the simulations did not change the intramolecular or *cis*-Ub–Ub interactions, they affected the character of intermolecular *trans*-Ub–Ub interactions (Extended Data Fig. 5a–d, left versus right). We also found that the membrane asymmetry enhanced the lifetime of the Ub–Ub interactions, stabilizing them to organize as a cluster (Extended Data Fig. 5e). Furthermore, whereas the intramolecular *cis*-interactions were predominantly mediated by specific residues on the Ub surface, namely, those forming hairpins β12, β34 and β5 (Extended Data Fig. 6a), the intermolecular *trans*-Ub–Ub interactions required for cluster formation were spread all over the Ub surface, indicating that no specific residues were dominant (Extended Data Fig. 6b). When the PPI strength was reduced to 65%, the alternative interactions that emerged

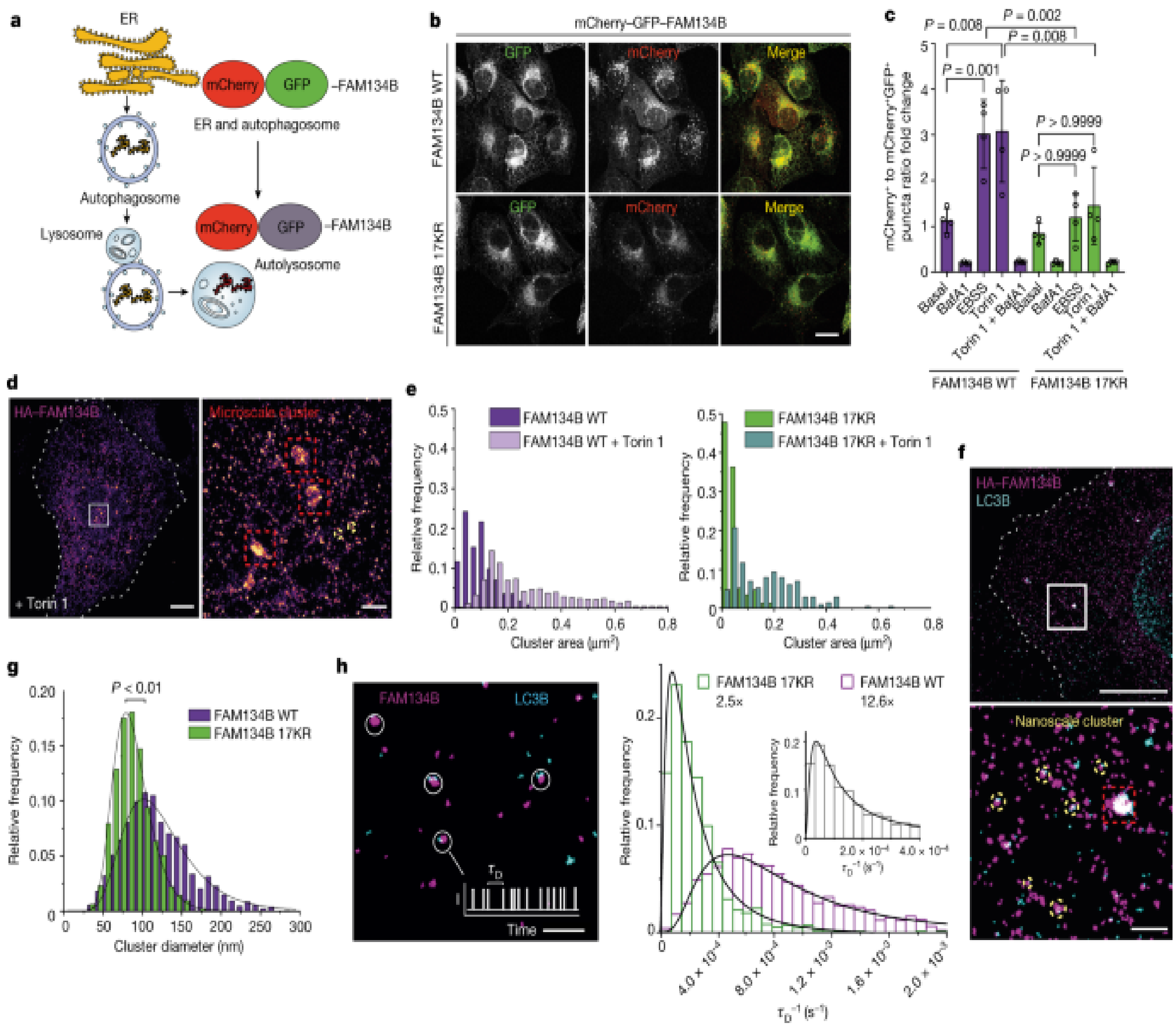
were also spread over the surface, confirming the absence of specific interactions. Nonspecific Ub–Ub interactions create a crowded membrane environment with multiple RHDs, causing proteins to sort and aggregate locally in the membrane. These crowded regions appear to be driven by volume-exclusion effects and curvature-mediated protein-sorting mechanisms. Thus, volume exclusion in combination with multiple low-affinity nonspecific Ub–Ub interactions result in the nucleation of RHD clusters, which then induce membrane bud formation. The high curvature of the membrane bud stabilizes the RHD clusters, increases their longevity and further favours the sorting of individual Ub–RHDs to the site of the bud.

RHD ubiquitination increases FAM134B cluster size and the flux of ER-phagy

Given that FAM134B functions in membrane remodelling, we hypothesized that ubiquitination affects FAM134B-driven ER fragmentation in cells. Using high-content imaging, we observed that the basal number and size of FAM134B–LC3B-decorated ER fragments were significantly lower in cells expressing the FAM134B 17KR mutant than in the WT control (Extended Data Fig. 7a–c). In vitro liposome remodelling driven by recombinant full-length GST-tagged FAM134B was not affected by the 17KR mutant, indicating that the folding and intrinsic activity of FAM134B were not impaired by the mutations (Extended Data Fig. 7d,e). Compared with liposomes treated with GST control, GST–FAM134B WT and 17KR decreased the liposome diameter to a similar extent (Extended Data Fig. 7f). Next, we investigated whether ubiquitination of the FAM134B RHD also affects the flux of ER-phagy using two validated reporter assays[3,7,8]. First, we generated cells expressing inducible mCherry–eGFP-tagged FAM134B 17KR or its WT counterpart (Fig. 3a). As expected, the reporter localized to the ER and the mCherry signal was concentrated in microscale puncta corresponding to autophagosomes or lysosomes (Fig. 3b). The lack of FAM134B RHD ubiquitination significantly reduced the flux of ER-phagy following Torin 1 or Earle's Balanced Salt Solution (EBSS) treatment (Fig. 3c). Similarly, the flux of ER-phagy was reduced when FAM134B WT was exposed to the E1 inhibitor TAK243 (Extended Data Fig. 7k). Second, in cells expressing the ER-phagy flux reporter ssRFP–GFP–KDEL[8], the flux of

ER-phagy (basal and induced by Torin 1 or EBSS) was reduced in cells co-expressing HA–FAM134B 17KR compared with the WT version (Extended Data Fig. 7g–j). Accordingly, less EBSS-induced REEP5 degradation was observed in the ssRFP–GFP–KDEL/HA–FAM134B 17KR cells (Extended Data Fig. 7j).

Fig. 3: Effect of RHD ubiquitination on the flux of ER-phagy and FAM134B cluster size.



a, The ER-phagy reporter system mCherry–GFP–FAM134B. **b**, U2OS TRex stable cell lines expressing mCherry–GFP–FAM134B WT or mCherry–GFP–FAM134B 17KR. Scale bar, 10 μm. **c**, ER-phagy flux was quantified as the ratio between mCherry$^+$GFP$^-$ and mCherry$^+$GFP$^+$ puncta. $n = 4$ independent experiments in which the total number of cells per condition for mCherry–GFP–FAM134B WT were: 482 (basal (DMSO)), 738 (BafA1),

673 (EBSS), 842 (Torin 1) and 667 (Torin 1 + BafA1). The number of cells per condition for mCherry–GFP–FAM134B 17KR: 440 (DMSO), 864 (BafA1), 723 (EBSS), 968 (Torin 1), 535 (Torin 1 + BafA1). Data are mean ± s.d.; one-way ANOVA, Bonferroni post-hoc test. **d**, DNA-PAINT super-resolution image of HA–FAM134B. Microscale (red) and nanoscale (yellow) clusters (red) are indicated. Scale bars, 10 μm (left panel) and 1 μm (right panel; magnified region from the left panel). **e**, Relative frequency distribution of HA–FAM134B WT and HA–FAM134B 17KR cluster areas ($110 > n_{cluster} < 251$) identified in U2OS cells under basal conditions (n_{WT} = 10 cells, n_{17KR} = 8 cells) or following Torin 1 treatment (n_{WT} = 11 cells, n_{17KR} = 14 cells). **f**, DNA-PAINT super-resolution imaging of HA–FAM134B WT and LC3B-II nanoscale clusters (yellow dashed circles). Dashed white line in top panel indicates cell outline. Scale bars, 10 μm (top panel) and 1 μm (bottom panel; magnified region from the top panel). **g**, Relative frequency distribution of the diameter of HA–FAM134B and HA–FAM134B 17KR nanoscale cluster ($n_{cells} = 4$, $n_{WT\ clusters}$ = 1,278; $n_{17KR\ clusters}$ = 1,255). Histograms were fitted with a log-normal distribution followed by a non-parametric one-tailed Mann–Whitney *U*-test. Cluster diameters were determined from the mode of the log-normal distribution using the mean and standard deviation (HA–FAM134B 17KR: $\mu = 88$nm, $\sigma = 24$ nm; HA–FAM134B WT: $\mu = 123$ nm, $\sigma = 46$ nm). **h**, Quantitative analysis of FAM134B copy numbers in nanoscale clusters. Scale bar, 1 μm. Relative frequency distribution of the inverse dark times (τ_D) of single-molecule binding time intervals recorded with DNA-PAINT (right). In the inset, the grey bars indicate calibration. Histograms were fitted with a log-normal distribution: FAM134B 17KR $\mu = 3.1 \times 10^{-4}\ s^{-1}$, $\sigma = 3.5 \times 10^{-4}\ s^{-1}$; FAM134B WT $\mu = 9.4 \times 10^{-4}\ s^{-1}$, $\sigma = 7.1 \times 10^{-4}\ s^{-1}$; calibration cluster $\mu = 1.9 \times 10^{-4}\ s^{-1}$, $\sigma = 2.8 \times 10^{-4}\ s^{-1}$.

Source data

We next investigated the ultrastructure of FAM134B clusters by single-molecule localization microscopy[9] in U2OS cells expressing HA–FAM134B WT or HA–FAM134B 17KR. Using DNA point accumulation in nanoscale topography (DNA-PAINT)[10], we identified microscale and nanoscale clusters of HA–FAM134B (Fig. 3d). The microscale clusters corresponded

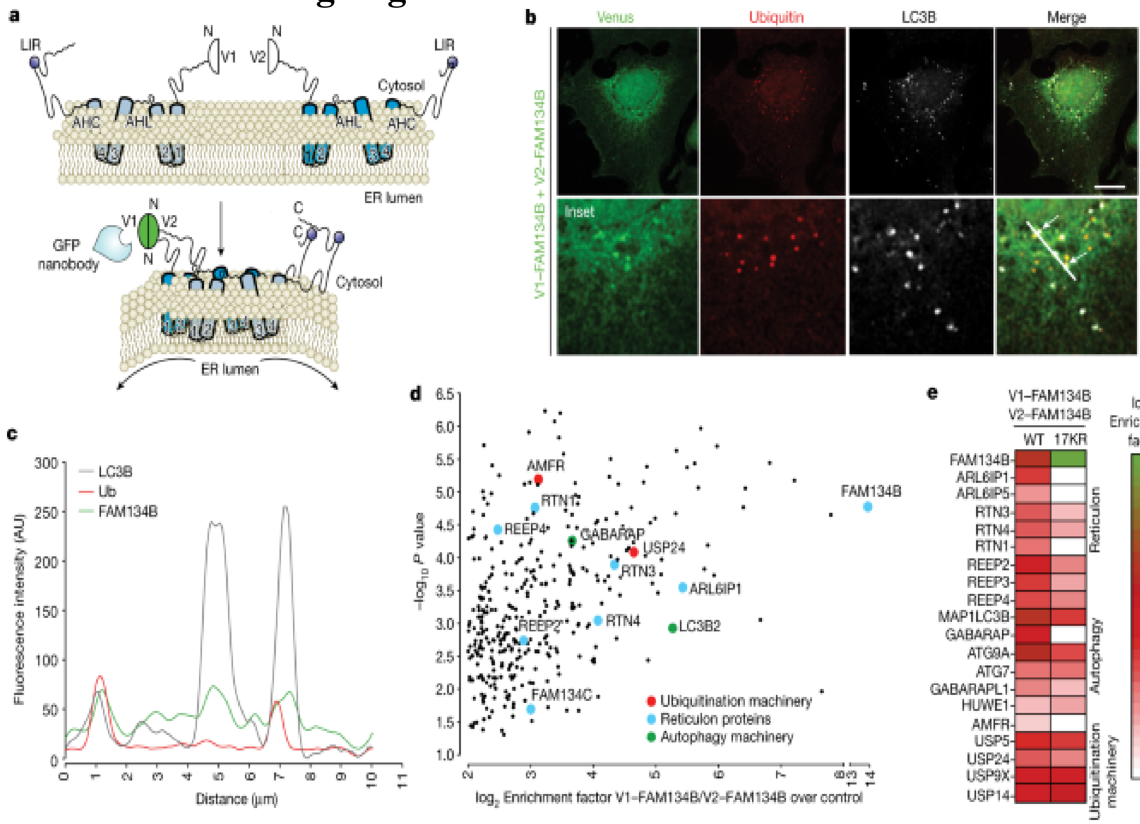
to the puncta observed by confocal microscopy (Extended Data Fig. 7a). Furthermore, 2D and 3D high-resolution images showed that microscale HA–FAM134B clusters colocalized with the autophagosomal membrane marker LC3B-II (Extended Data Fig. 7l and Supplementary Video 7). We determined the size of HA–FAM134B WT and HA–FAM134B 17KR microscale clusters by Voronoi tessellation[11], indicating the area for the heterogeneous morphologies that we observed. Torin 1 treatment resulted in larger clusters for HA–FAM134B WT and the ubiquitination-deficient mutant (Fig. 3e). However, the cluster areas were significantly larger for HA–FAM134B WT than the mutant under control conditions and following exposure to Torin 1 (Fig. 3e and Extended Data Fig. 7m). Super-resolution images (28 ± 2 nm, mean ± s.d.)[11] revealed the existence of HA–FAM134B nanoclusters much smaller than the resolution of light microscopy (Fig. 3d, dotted dashed circle). Co-labelling with REEP5 revealed that HA–FAM134B nanoscale clusters were distributed within the ER network (Extended Data Fig. 7n). Using DNA-PAINT super-resolution imaging, we found that a subset of HA–FAM134B nanoscale clusters colocalized with the autophagosomal membrane marker LC3B-II, and may represent ER-phagy initiation sites (Fig. 3f). Quantitative analysis of these sites using the DBSCAN clustering algorithm[12] revealed that the diameter of the clusters containing ubiquitination-deficient HA–FAM134B was significantly lower (diameter = 79 nm) than those containing FAM134B WT (diameter = 101 nm) (Fig. 3g). We inferred the number of molecules in the nanoscale clusters by applying a kinetic analysis of single-molecule DNA-PAINT data[13], revealing that ubiquitination increased the oligomeric state of FAM134B WT in nanoclusters on average by fivefold ($n_{17KR} = 2.5$, $n_{WT} = 12.6$), thus promoting the assembly of high-density clusters (Fig. 3h). These data suggest that FAM134B RHD ubiquitination increases the size of ER-phagy initiation clusters, leading to larger autophagosomal structures that tune the dynamic flux of ER-phagy.

Protein interactors of ubiquitinated FAM134B clusters in cells

To investigate the composition of the ER-phagy receptor complexes in more detail, we used a bimolecular complementation affinity purification (BiCAP)

assay to visualize the PPIs and to characterize the cluster-specific interactome[14] (Fig. 4a). The clustering of FAM134B was enriched by immunoprecipitation with anti-GFP antibodies following the co-expression of V1–FAM134B and V2–FAM134B (Extended Data Fig. 8a, compare lane 3 to lanes 2 and 1). Similar results were observed following the BiCAP of FAM134C, a paralogue that acts in concert with FAM134B[5] (Extended Data Fig. 8a, compare lane 7 to lanes 6 and 5). The analysis of diGly peptides from isolated FAM134B oligomers revealed seven ubiquitinated lysine residues within the FAM134B RHD, including the four previously detected sites (Fig. 1b and Extended Data Fig. 8b). This indicated that multiple lysine residues in the RHD can be ubiquitinated within receptor clusters. Accordingly, FAM134B clusters (observed using BiCAP) were found in ER fragments, colocalizing with LC3B and Ub (Fig. 4b,c), indicating that ubiquitinated FAM134B clusters are colocalized with $LC3^+$ autophagosomes. We characterized the interactome of immunopurified FAM134B clusters and detected 363 significantly enriched proteins ($\log_2$ enrichment factor ≥ 2, $P \leq 0.05$), including novel candidates that had not been detected in previous FAM134B interactome datasets[15]. MAP1LC3B and GABARAP were identified among the most enriched proteins (Fig. 4d, green dots). In addition, we found that FAM134B homodimers strongly interacted with several other RHDs and RHD-containing proteins (Fig. 4d, light blue dots, and Extended Data Fig. 8g) and with proteins of the ubiquitination machinery (Fig. 4d, red dots), including E3 ligases and deubiquitinases. Autophagy receptors containing Ub-binding domains (for example, p62, OPTN and TAX1BP1)[16] were not detected, implying that FAM134B ubiquitination is not a recruitment signal for these proteins. Accordingly, there was no significant colocalization between FAM134B clusters and p62 in cells (Extended Data Fig. 8h,i). The interactomes of FAM134 isoforms, FAM134C heterodimers and FAM134B–FAM134C dimers revealed overlapping sets of interaction partners (Extended Data Fig. 8c–f). By contrast, the clustering of FAM134B 17KR reduced or abolished interactions between FAM134B and several RHD-containing proteins, mammalian ATG8 proteins and the ubiquitination machinery (Fig. 4e), indicating that ubiquitination promotes the formation of multimeric ER-phagy clusters.

Fig. 4: Ubiquitination site profiling and protein interactors of FAM134B-containing oligomers.



a, Schematic representation of the bimolecular complementation affinity purification assay of FAM134B dimers. The full-length FAM134B was fused to the C-terminal of the two non-fluorescent complementary fragments of the Venus fluorescent protein (V1–FAM134B and V2–FAM134B). **b**, Confocal fluorescence microscopy analysis of the bimolecular fluorescence complementation signal resulting from the interaction between V1–FAM134B and V2–FAM134B. Fixed cells expressing V1–FAM134B and V2–FAM134B were stained with anti-Ub(FK2) (red) or anti-LC3B (grey). Arrows indicate triple colocalization of the Venus signal (green), ubiquitin (red) and LC3B (grey). Scale bar, 10 μm. **c**, Histogram analysis of the fluorescence intensity distribution reveals that dimeric FAM134B colocalized into Ub^+ and $LC3B^+$ vesicles. **d**, Single-sided volcano plot of the quantitative label-free interactome of FAM134B homodimers depicting RHD-containing ER proteins (blue), autophagy-related proteins (green) and ubiquitination machinery (red). Data represent three independent experiments, one-sided unpaired Student's *t*-test. **e**, Heatmap comparing the interaction of FAM134B WT homodimers versus FAM134 17KR

homodimers with RHD-containing ER proteins, autophagy-related proteins and ubiquitination machinery. Interaction partners with $\log_2$ enrichment > 2.0 and $-\log_{10} P > 1.3$ were plotted and compared (one-sided unpaired Student's *t*-test).

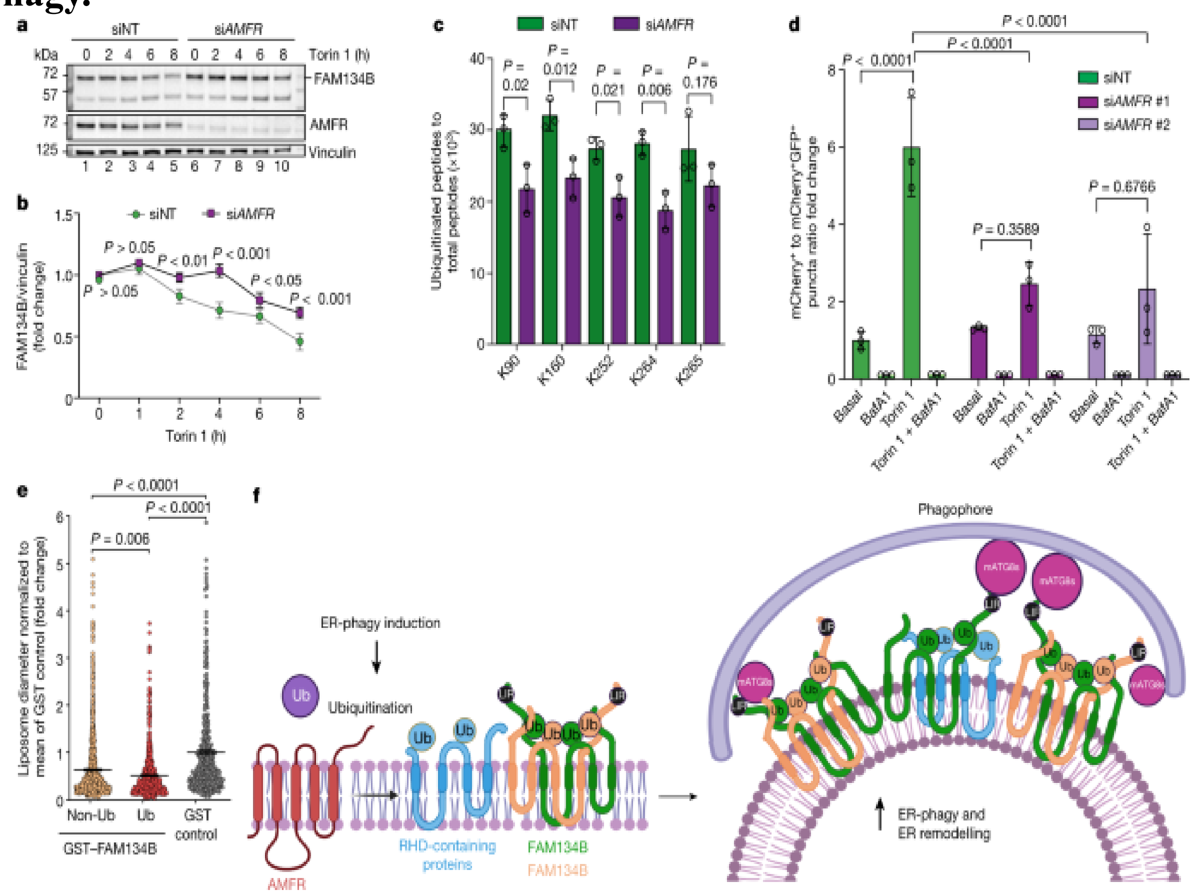
Source data

The E3 ubiquitin ligase AMFR catalyses FAM134B ubiquitination

The interaction partners within oligomeric FAM134B clusters included the endogenous ER-anchored E3 ligase AMFR (also known as gp78), which is implicated in ER-associated degradation[17] (Fig. 4d). This colocalized with FAM134B clusters in BiCAP experiments and with LC3B (Extended Data Fig. 9a,b). The expression of WT AMFR, but not its catalytically inactive counterpart (C356G H361A), increased the ubiquitination of HA–FAM134B in cells (Extended Data Fig. 9c,d). Evaluating their functional interaction in cells using BiCAP assays (Extended Data Fig. 9e) revealed that V2–AMFR–V1–FAM134B complexes colocalized with LC3B and Ub^+ structures (Extended Data Fig. 9f). This effect was reduced when AMFR was replaced with its catalytically inactive mutant (Extended Data Fig. 9f,g). MS-based interactome analysis confirmed that FAM134B–AMFR complexes also recruited RHD-containing proteins, mammalian ATG8 proteins and components of the ubiquitination machinery (Extended Data Fig. 9h,i). Of note, the E2-conjugating enzyme UBE2G2, which cooperates with AMFR[18], was also enriched (Extended Data Fig. 9h). The Torin 1-induced ubiquitination profile of FAM134B complexes with AMFR (WT V2–AMFR–V1–FAM134B) was significantly reduced in complexes with the catalytically inactive AMFR mutant (V2–AMFR(C356G,H361A)–V1–FAM134B) (Extended Data Fig. 9j). In addition, clustering of FAM134B with catalytically inactive AMFR reduced the interactions with other RHD-containing proteins, mammalian ATG8 proteins and the ubiquitination machinery (Extended Data Fig. 9k). Silencing the expression of AMFR caused a significant increase in total endogenous FAM134B (Extended Data Fig. 10a,b) and diminished its cellular turnover induced by Torin 1 (Fig. 5a,b). Furthermore, we observed decreased ubiquitination of S protein-

FLAG-streptavidin-binding peptide (SFB)-tagged FAM134B in response to AMFR depletion (Fig. 5c), also reducing FAM134B-mediated ER fragmentation in response to Torin 1 (Extended Data Fig. 10c,d). Torin 1 treatment also reduced the levels of endogenous AMFR (Fig. 5a and Extended Data Fig. 10e). This decay was diminished in the presence of HA–FAM134B 17KR (Extended Data Fig. 10f–h) or ΔLIR (Extended Data Fig. 10i,j), indicating that active FAM134B and its ubiquitination promote efficient AMFR degradation via ER-phagy. Moreover, siRNA-mediated knockdown of *AMFR* significantly slowed the flux of FAM134B-mediated Torin 1-induced ER-phagy compared with control siRNA (siNT) (Fig. 5d; approximately 58% and 60% reduction with *AMFR*-targeting siRNA #1 (si*AMFR* #1) and si*AMFR* #2, respectively).

Fig. 5: Effect of AMFR on FAM134B RHD ubiquitination and ER-phagy.



a, HeLa cells were transfected with siNT or si*AMFR*, and were treated with 250 nM Torin 1 for the indicated time. Protein extracts were analysed by western blot for FAM134B, AMFR or vinculin. **b**, Densitometric

quantification of the western blot signal of FAM134B in **a** (data are mean ± s.d.; $n = 3$ independent experiments; two-way ANOVA, Bonferroni post-hoc test). **c**, Ubiquitination of SFB-tagged FAM134B in *AMFR*-knockdown cells. The diGly peptide intensities were normalized to total intensities of modified and non-modified FAM134B peptides (data are mean ± s.d.; $n = 3$ independent experiments; two-way ANOVA, Bonferroni post-hoc test). **d**, U2OS TRex stable cell lines expressing mCherry–GFP–FAM134B WT were transfected with siNT, *AMFR*-targeting siRNA #1 or siRNA #2. Cells were treated as indicated. The flux of ER-phagy was quantified as the ratio between mCherry^{+}GFP^{-} and mCherry^{+}GFP^{+} puncta. The data are representative of three independent experiments in which the total number of cells per condition were: 699 siNT/DMSO, 683 siNT/BafA1, 695 siNT/Torin 1, 678 siNT/Torin 1 + BafA1, 627 si*AMFR* #1/DMSO, 668 si*AMFR* #1/BafA1, 713 si*AMFR* #1/Torin 1, 591 si*AMFR* #1/Torin 1 + BAfA1, 624 si*AMFR* #2/DMSO, 593 si*AMFR* #2/BafA1, 575 si*AMFR* #2/Torin 1 and 615 si*AMFR* #2/Torin 1 + BafA1. Data are mean ± s.d., one-way ANOVA, Bonferroni post-hoc test. **e**, Diameters of freeze-fractured liposomes incubated with either non-ubiquitinated or ubiquitinated GST–FAM134B. Data are mean ± s.e.m. normalized to the mean liposome diameter of GST control, representing three independent liposome preparations and experiments ($n = 1,064$ for the non-Ub sample, $n = 793$ for the Ub sample and $n = 434$ for the GST control; Kruskal–Wallis/Dunn's post-test). **f**, Model in which the E3 ligase AMFR is recruited to ER-phagy receptor clusters to induce the ubiquitination of FAM134B. This event triggers changes in the conformation and composition of ER-phagy receptor clusters, enabling the clusters to grow in size, thus controlling ER remodelling and ER-phagy. mATG8s, mammalian ATG8 proteins. The schematic in panel **f** was created using BioRender (https://biorender.com).

Source data

Next, we tested whether AMFR-mediated ubiquitination affected the membrane-shaping functions of FAM134B in vitro. First, we ubiquitinated GST-tagged FAM134B using purified recombinant AMFR (Extended Data Fig. 10k,l) and detected K160, K278 and K299 as direct targets of AMFR-dependent ubiquitination (Extended Data Fig. 10m,n). This prompted us to repeat the in vitro liposome remodelling assay in the presence of AMFR. Compared with liposomes treated with GST, the non-ubiquitinated GST–

FAM134B (in the presence of AMFR, no ATP) decreased the liposome diameter, but ubiquitinated GST–FAM134B (in the presence of AMFR + ATP) reduced the diameter even further (Fig. 5e). This significant difference between non-ubiquitinated and ubiquitinated samples suggests that ubiquitination of multiple sites on full-size FAM134B promotes the formation of smaller liposomes. This agrees with the increased membrane remodelling activity for the chimaera Ub–$RHD_{90–264}$–Ub compared with non-ubiquitinated $RHD_{90–264}$ (Fig. 2c). Furthermore, the ubiquitination of FAM134B $RHD_{90–264}$ in vitro by AMFR promoted the formation of larger complexes detected by blue native polyacrylamide gel electrophoresis (Extended Data Fig. 10o–r).

Discussion

Our data reveal an unprecedented role for Ub in the conformational changes of FAM134B that, in turn, drive ER membrane remodelling and the flux of ER-phagy. Although FAM134B RHDs inherently curve bilayers and populate regions of high local membrane curvature, the addition of Ub boosts this effect by promoting multiple *cis*-interactions and *trans*-interactions that stabilize multimeric clusters. This enables the clusters to grow in size and to nucleate large-scale membrane remodelling events in the ER (Fig. 5f). ER-phagy clusters also include multiple RHD-containing proteins such as ARL6IP1 (a recurrent finding in the interactome of FAM134B homodimers), which is also required for efficient ER-phagy[19]. In addition, the clusters contain Ub ligases and deubiquitinases, which can alter the dynamic ubiquitination of RHD-containing proteins and influence the formation and growth of these multivalent clusters. Indeed, the E3 ligase AMFR regulates ER-phagy by acting as a critical ER quality control mechanism. It is tempting to speculate that other E3 ligases can modify ER-phagy receptor clusters in a cell-type-specific manner. The receptor clustering phenomenon depends on several interrelated factors, such as receptor abundance, distribution and other post-translational modifications (for example, phosphorylation[7] or UFMylation[20]). Studying these events will shed light on the dynamics of the entire ER-phagy pathway and will pave the way for a better understanding of defects in ER dynamics influencing the pathogenesis of many diseases[21].

Methods

Cell culture and inducible cell lines

Cell lines were cultured in DMEM supplemented with 10% heat-inactivated FBS and penicillin–streptomycin (Thermo Fisher Scientific) in a 5% CO_2 atmosphere at 37 °C. U2OS TRex cells (provided by S. Blacklow, Brigham and Women's Hospital and Harvard Medical School) and U2OS cell lines (American Type Culture Collection (ATCC)) were used to generate inducible cell lines based on lentiviral infection[22]. The mCherry–eGFP–FAM134B WT and mCherry–eGFP–FAM134B 17K constructs were introduced into the vector pcDNA5/FRT/TO using GATEWAY technology and transfected with the recombinase vector pOG44 into Flp-In U2OS TRex cells. Following selection with 300 μg ml^{-1} hygromycin, the resistant cells were expanded. We generated U2OS cell lines expressing HA–FAM134B or HA–FAM134B 17KR under the control of a doxycycline-inducible promoter. Expression was induced with 0.5 μg ml^{-1} doxycycline for the indicated time. Lentiviruses were produced in HEK 293T cells (ATCC). In brief, HEK 293T (ATCC) cells were co-transfected with the lentiviral plasmids (1.1 μg complementary DNA (cDNA)), containing the cDNA of FAM134B WT or 17KR along with the two packaging vectors pPAX2 (2.2 μg cDNA) and pMD2.G (1 μg cDNA) using Turbofect reagent (Thermo Fisher Scientific). The lentivirus-containing medium was collected after 48 h. The medium was centrifuged to remove dead HEK 293T cells and stored at −80 °C. After 24 h, cells were selected in fresh DMEM containing 3 μg ml^{-1} puromycin. Cells were maintained in the presence of antibiotics for selection and were grown to subconfluence before each experiment. Cells were treated with 0.5 μg ml^{-1} doxycycline (Sigma-Aldrich), 200 ng ml^{-1} bafilomycin A1 (LC Laboratories), 50 μg ml^{-1} cycloheximide (AppliChem PanReac), 250 nM Torin 1 (LC Laboratories) and/or 1 μM or 10 μM TAK243 (MedChemExpress) for the indicated time periods. Transient transfection was carried out using Turbofect reagent. All cell lines were regularly tested for mycoplasma contamination using the LookOut Mycoplasma PCR Detection Kit (Sigma).

Human *AMFR*-targeting siRNAs (siRNA #1 5′-GCA AGG AUC GAU UUG AAU A-3′; and siRNA #2, 5′-GUA AAU ACC GCU UGC UGU G-3′) were purchased from Dharmacon. A non-targeting siRNA was used as a control (Qiagen). Transfection with siRNA was carried out using Lipofectamine RNAiMAX (Thermo Fisher Scientific). Assays were carried out 72 h post-transfection.

Antibodies

Primary and secondary antibodies are presented in Supplementary Table 1.

Plasmids

We introduced the cDNAs into vector pDONR223 using the BP Clonase Reaction Kit (Invitrogen) followed by transfer to GATEWAY destination vectors using the LR Clonase Reaction Kit (Invitrogen), resulting in vectors pLTD-N-HA-PURO, pcDNA5-FRT/TO-N-mCherry-eGFP, pcDNA3.1-N-HA, pMH-SFB, pGEX6P1-DEST, pDEST 527 6×His, pDEST-V1ORF, pDEST-V2-ORF, pDEST-ORF-V1 and pDEST-ORF-V2. GST–FAM134B was generated by subcloning into the vector pGEX-6P1 using the EcoRI site. GST–FAM134B 17KR was similarly generated by targeting the EcoRV and SmaI sites of pGEX-6P2.

Plasmids are presented in Supplementary Table 2.

Ubiquitination assays in cells, co-immunoprecipitation and TUBE-2 pulldown

To assess the ubiquitination status of FAM134B, HEK 293T cells transfected with Myc–Ub, HA–FAM134B constructs and WT AMFR–Flag or its catalytically inactive mutant as indicated were lysed (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 1% Triton X-100, 10 mM *N*-ethylmaleimide (NEM) and Roche EDTA-free protease inhibitors, freshly added). The lysates were incubated for 15 min on ice and centrifuged (12,000*g* at 4 °C for 30 min) before 20 µl of the supernatant was supplemented with Laemmli sample buffer, boiled for 5 min at 95 °C and stored at –20 °C as the input control. Ubiquitinated proteins

(Myc–Ub proteins) were immunoprecipitated from soluble extracts with Myc-Trap agarose (Chromotek). Beads were washed three times in lysis buffer and proteins were denatured by heating at 95 °C for 5 min before SDS–PAGE and western blot analysis with anti-HA antibodies for the detection of HA–FAM134B constructs. For other immunoprecipitation assays, cleared lysates were incubated with GFP-Trap (Chromotek), HA-agarose beads (Sigma-Aldrich) or TUBE-2 agarose beads (Life Sensors) at 4 °C overnight. The next day, tubes were centrifuged (800 rcf at 4 °C for 5 min) to sediment the beads, and the supernatant was removed and washed with ice-cold lysis buffer. The inputs and co-precipitation fractions were analysed by SDS–PAGE and western blot.

Blue-native PAGE, SDS–PAGE and western blot

Blue-native PAGE was performed following the manufacturer's instructions (#BN1001BOX, protocol pub. no. MAN0007893 Rev. A.0, Life Technology). For SDS–PAGE, proteins were denatured at 90 °C for 5 min in Laemmli buffer. Native or denatured proteins were transferred to methanol-activated polyvinylidene fluoride membranes (Amersham Hybond P, 0.45 μm). Membranes were blocked for 1 h in 5% skimmed milk in PBS containing 0.1% Tween-20 and incubated overnight at 4 °C with the specified primary antibody, and for 2 h with the corresponding secondary antibody at room temperature. We used horseradish peroxidase-conjugated anti-rabbit (1:10,000), anti-mouse (1:10,000) and anti-rat (1:10,000) secondary antibodies as appropriate, followed by signal development using Western ECL substrate (sc-2048, SantaCruz) and the Chemidoc automated detection system (Bio-Rad). Densitometric quantification of western blot bands was carried out using ImageJ (version 1.51w).

Densitometric quantification and statistical analysis

Quantifications of western blot signals were performed using ImageJ software (version 1.51w). For each assay, protein bands were quantified from at least three independent experiments. Data analysis was performed using Microsoft Excel 2016 (Microsoft Corporation) or Prism 9.4.1 (GraphPad Software). Results were graphed as the mean ± standard deviation. Statistical significance was determined by one-tailed, paired *t*-test.

$P > 0.05$ or $P \leq 0.05$ were regarded as not statistically significant or statistically significant, respectively. One-way analysis of variance (ANOVA) was used in assays with two independent variables. Bonferroni's multiple comparison test was performed.

Fluorescence microscopy

Cells were grown on 12-mm glass coverslips or in 96-well plates. Cells were washed with PBS before fixation in 4% paraformaldehyde (PFA) for 10 min at room temperature. After several further washes with PBS, cells were permeabilized with 0.25% saponin in PBS and blocked with 5% FBS for 1 h at room temperature. Primary antibodies were diluted in 5% FBS/0.25% saponin in PBS and incubated overnight at 4 °C. After three PBS washes, secondary antibodies were added and incubated for 1 h at room temperature. The cells were washed another three times before staining with the nuclear dye DAPI for 10 min, and washed again before mounting with Fluoromount-G. The following secondary antibodies were used in a 1:1,000 dilution: anti-rabbit Alexa 488, anti-mouse Alexa 546 and anti-rat Cy5. Slides were imaged using a Leica SP8 confocal microscope fitted with a ×63 oil-immersion lens and analysed with ImageJ software (version 1.51w). For ER fragmentation and ER-phagy flux assays, doxycycline-inducible mCherry–GFP–FAM134B U2OS cells were analysed using a Yokogawa CQ1 (vR1.08.01) confocal imaging cytometer.

Protein expression and purification

Escherichia coli (BL21(DE3)) cells were transformed with constructs encoding GST-tagged FAM134B (WT, 17KR mutant, $RHD_{90–264}$ or Ub–$RHD_{90–264}$–Ub), GST–Ub and GST, followed by expression and purification as previously described[4,23]. In brief, transformed cells were grown in the presence of appropriate antibiotics overnight and the primary culture was used to inoculate the main culture. When the $OD_{600\ nm}$ reached 0.6, 0.25 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) was used to induce protein expression for 16 h at 18 °C, shaking at 180 rpm. After induction, the cells were harvested and the pellets were suspended in ice-cold PBS, followed by sonication and centrifugation (10,000*g* for 30 min at 4 °C). The supernatant was then fractionated by ultracentrifugation (80,000*g* for 90 min at 4 °C).

RHD fusion proteins were recovered by dissolving the pellets in PBS containing 0.05% dodecyl β-d-maltoside (DDM) and loading onto glutathione-Sepharose TM4 Fast Flow columns (GE Healthcare). The columns were washed with PBS containing 0.05% DDM and the proteins were eluted in PBS containing 15 mM reduced glutathione and 0.025% DDM. GST–Ub and GST were directly loaded from the centrifugation step after sonication and were purified on the GST column. The eluted fractions were concentrated and exchanged with storage buffer (50 mM HEPES pH 7.5, 150 mM NaCl and 0.0075% DDM).

Alternatively, cells expressing GST–FAM134B WT and 17KR were lysed in a French press (G. Heinemann Ultraschall & Labortechnik) to avoid protein denaturation by sonication. Bacterial lysates were mixed with 2% (v/v) Triton X-100, 20 mM $MgCl_2$, 40 μg ml^{-1} DNAse I and 4 mg ml^{-1} lysozyme for 30 min at 1 °C, and cell debris was removed by centrifugation (1,000*g* for 20 min at 4 °C). The supernatant was incubated with ice-cold PBS, and GST fusion proteins were affinity purified on glutathione-Sepharose resin (Genscript). After washing, bound proteins were eluted by incubation with 20 mM glutathione in 50 mM Tris-HCl pH 8.0 and 120 mM NaCl. The purified proteins were concentrated using Amicon Ultra-4-10k centrifugal filters (Millipore) and dialysed at 4 °C against HN-buffer (20 mM HEPES/KOH pH 7.4, 150 mM NaCl and 2.5 mM DTT) for liposome and freeze-fracture assays.

For the purification of His–$RHD_{90–264}$–Strept-II and His–Ub–$RHD_{90–264}$–Ub–Strept-II, bacterial pellets were resuspended in ice-cold binding buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM TCEP) plus a cocktail of protease inhibitors (1 mM PMSF, 1 μM GM6001, 0.25 μM bestatin, 0.5 μM pepstatin and 1 μM E-64), DNAse I (50 μg ml^{-1}) and 0.1% DDM. The cells were disrupted by two passes through a microfluidizer at 1,500 bar, and the debris was removed by centrifugation (12,000*g* for 1 h at 4 °C). The supernatant, including the membranes, was centrifuged (43,000*g* for 2.5 h at 4 °C) on an Ultracentrifuge Optima L-90K with a 45 Ti rotor (Beckman Coulter). The pelleted membrane was solubilized in membrane extraction buffer (100 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM EDTA, 25% glycerol and 2% DDM) at 4 °C with gentle stirring for approximately 2 h. The insoluble fractions were removed by ultracentrifugation (55,000*g* for

1 h at 4 °C) and the supernatant containing the solubilized membrane proteins was passed through a 0.22-μM filter and supplemented with 40 μg ml^{-1} avidin before purification. The first purification step targeted the strep II tag. A Hiprep strep II 5-ml column (Cytiva) on an Äkta FPLC system was pre-equilibrated with binding buffer supplemented with 0.1% DDM. The sample was allowed to bind to the column and then washed with binding buffer containing 0.01% DDM and lacking EDTA. Fractions in elution buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.03% DDM and 3 mM d-desthiobiotin) were analysed by SDS–PAGE and western blot with antibodies specific for the strep II tag. The desired fractions were pooled and supplemented with 25 mM imidazole followed by second-step purification targeting the His_6 tag. Talon beads were used and a linear gradient between buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 25 mM imidazole, 5% glycerol, 0.03% DDM and 1 mM TCEP) and buffer B (buffer A with 250 mM imidazole) was used to purify the protein. The resulting protein was buffer exchanged (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM TCEP and 0.03% DDM) using a centricon filter with a 10-kDa cut-off. The protein was aliquoted and flash frozen in liquid nitrogen for further experiments.

For AMFR expression in mammalian cells, 3 l cultures of suspension-adapted HEK 293T cells were grown in customized DMEM[24] to a density of 1×10^6 cells per millilitre. We added 15% (v/v) P3 baculovirus to each flask and incubated on a shaking platform for 24 h, after which the temperature was reduced to 30 °C and 10 mM sodium butyrate was added. Cells were harvested after shaking for 48 h and the pellets were frozen at –80 °C. For AMFR protein purification, approximately 15 g of cells was thawed on ice and resuspended in 60 ml lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM DTT, Roche protease inhibitor cocktail, 1 mM $MgCl_2$ and 0.001 mg ml^{-1} Benzonase) and sonicated on ice. To solubilize membranes, DDM and CHS was added to the lysate at a final concentration of 1% and 0.1%, respectively, and stirred for 1 h at 4 °C. Insoluble material was removed by centrifugation (40,000*g* for 30 min at 4 °C). The supernatant was incubated with 2 ml StrepTactin resin (Cytiva) for 2 h at 4 °C to bind to gp78-TEV-TwinStrepII. The resin was poured into a column and washed with 24 column volumes (CV) of wash buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM DTT and 0.05/0.005% DDM/CHS). Bound protein was

eluted with wash buffer containing 5 mM desthiobiotin. To remove the C-terminal affinity tag, eluted protein was incubated for 2 h with 0.1 mg ml^{-1} TEV protease at room temperature. Finally, pooled elution fractions were concentrated to 0.5 ml using 100-kDa cut-off centrifugal concentrators (Amicon) and purified by gel filtration on a Superose 6 10/300 column (Cytiva) running with SEC buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 5 mM DTT and 0.05/0.005% DDM/CHS).

Liposome preparation, liposome shaping assay and electron microscopy

Liposomes made from synthetic lipids were prepared as previously described[4]. In brief, 2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and 1,2-dioleoyl-*sn*-glycero-3 phosphoethanolamine (DOPE), both from Avanti Polar Lipids, were dissolved in a mixture of chloroform and methanol (4:1) in a round-bottom flask at a molar ratio of 0.8:0.2 (DOPC:DOPE). The organic solvent was removed by rotary evaporation to obtain a uniform dry lipid film, which was then hydrated for 2 h at room temperature with liposome buffer A (50 mM HEPES pH 7.4 and 150 mM NaCl) to obtain a final 15 mg ml^{-1} solution. Liposomes were dissolved by vortexing followed by sonication in an ultrasound bath. Liposomes were equilibrated to 25 °C and extruded using a lipid extruder with 200-nm polycarbonate membranes (Avanti Polar Lipids).

For negative staining assays, 2.5 µM FAM134B WT and FAM134B 17KR mutant or GST were incubated with 1 mg ml^{-1} liposomes in liposome buffer B (50 mM HEPES pH 7.4, 150 mM NaCl and 0.001% DDM) for 18 h at 25 °C on a table-top shaker at 600 rpm. GST–FAM134B $RHD_{90–264}$ or GST–Ub–$RHD_{90–264}$–Ub chimaera and Ub–GST were incubated at 0.5 µM with 1 mg ml^{-1} liposomes in liposome buffer B for 8 h. We then added 5 µl of each sample to the carbon-coated copper grids (SPI Supplies) without glow discharge. After 1 min, the grids were washed twice with water and stained with 1% uranyl formate for 1 min at room temperature. Excess solution was removed by blotting with filter paper. Approximately 20 micrographs were recorded for each sample using a 120 kV Tecnai Spirit Biotwin electron microscope (FEI) equipped with a 4k × 4k CCD detector (US4000-1, Gatan).

Alternatively, membrane shaping by full-length GST–FAM134B ubiquitinated with AMFR was investigated by transmission electron microscopy using freeze-fractured liposomes prepared from Folch-fraction type I lipids (Sigma-Aldrich) as previously described[25]. We incubated 1 mg of liposomes with 2.5 µM protein in HN-buffer (20 mM HEPES/KOH pH 7.4, 150 mM NaCl and 2.5 mM DTT) containing 0.3 M sucrose for 15 min at 37 °C. Subsequently, 15 µg proteinase K was added and incubated for 40 min at 45 °C to avoid liposomal aggregates[26]. Small aliquots (1–2 µl) of the liposome suspension were then freeze-fractured[26,27]. The samples were examined by systematic grid exploration using an EM 900 electron microscope (Zeiss) at 80 kV. Images were acquired using a wide-angle dual-speed 2K CCD camera (Tröndle). Diameters of liposomes were determined using ImageJ software (version 1.51w).

In vitro ubiquitination of FAM134B using recombinant AMFR

AMFR-mediated ubiquitination assays were based on a modified in vitro ubiquitination assay (Abcam). In brief, purified substrate (5 µM full-length GST–FAM134B, 1 µM His–$RHD_{90–264}$–Strept-II or 1 µM His–Ub–$RHD_{90–264}$–Ub–Strept-II) was mixed with 10 µM Ub, 10 mM ATP and 10 mM $MgCl_2$ in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.8 µM E3-ligase AMFR, 100 nM E1 UBA1 and 0.8 µM E2 of AMFR UBE2G2 (Biotechne) for 2 h at 37 °C. The reaction mixture was analysed by SDS–PAGE or western blotting with antibodies against GST, His_6 or Ub(P4D1). Alternatively, the samples were prepared for MS. Samples were incubated with SDC buffer (1% sodium deoxycholate, 0.5 mM TCEP, 2 mM chloroacetamide and 50 mM Tris-HCl pH 8.5) and heated to 60 °C for 30 min. We then added 500 ng trypsin to each sample and incubated overnight at 37 °C. The reaction was stopped with 1% TFA in isopropanol. Peptides were cleaned up using SDB-RPS stage tips (Sigma-Aldrich). After one wash with 1% TFA in isopropanol and one wash with 0.2% TFA in water, peptides were eluted in 80% acetonitrile plus 1.25% ammonia. Eluted peptides were dried and processed for LC–MS.

BiCAP interactome analysis and sample preparation for MS

HEK 293T cells were transiently co-transfected with the constructs V1–FAM134B WT and V2–FAM134B WT, V1–FAM134C WT and V2–FAM134C WT, V1–FAM134B WT and V2–FAM134C or V1–FAM134B 17KR and V2–FAM134B 17KR. After 16 h, cells were lysed with 1% Triton X-100 in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, Roche EDTA-free protease inhibitor cocktail and NEM), followed by incubation with GFP-trap beads (Chromotek) overnight at 4 °C on a rotating platform. Protein-bound beads were washed three times with lysis buffer and three times in the same buffer without detergents before on-bead trypsin digestion. Samples were incubated with 25 µl SDC buffer (2% sodium deoxycholate, 1 mM TCEP, 4 mM chloroacetamide and 50 mM Tris-HCl pH 8.5) and heated to 60 °C for 30 min. We then added 500 ng trypsin in 25 µl 50 mM Tris-HCl (pH 8.5) to each sample and incubated overnight at 37 °C. The reaction was stopped with 150 µl of 1% TFA in isopropanol. Peptides were cleaned up using SDB-RPS stage tips (Sigma-Aldrich). After one wash with 1% TFA in isopropanol and one wash with 0.2% TFA in water, peptides were eluted in 80% acetonitrile plus 1.25% ammonia. Eluted peptides were dried and processed for LC–MS.

LC–MS analysis

Dried peptides were reconstituted in 2% acetonitrile containing 0.1% TFA and analysed on a QExactive HF mass spectrometer coupled to an easy nLC 1200 (Thermo Fisher Scientific) fitted with a 35-cm, 75-µm ID fused-silica column packed in house with 1.9-µm C18 particles (Reprosil pur, Dr. Maisch). The column was maintained at 50 °C using an integrated column oven (Sonation). Peptides were eluted in a non-linear gradient of 4–28% acetonitrile over 45 min and directly sprayed into the mass spectrometer equipped with a nanoFlex ion source (Thermo Fisher Scientific). Full-scan MS spectra (300–1,650 *m*/*z*) were acquired in profile mode at a resolution of 60,000 at *m*/*z* 200, a maximum injection time of 20 ms and an AGC (automatic gain control) target value of 3×10^6. Up to 15 of the most intense peptides per full scan were isolated using a 1.4-Th window for fragmentation by higher energy collisional dissociation (normalized collision energy of 28). MS/MS spectra were acquired in centroid mode with a resolution of 30,000, a maximum injection time of 45 ms and an AGC target value of 1×10^5. Single charged ions, ions with a charge state of more than

four and ions with unassigned charge states were not considered for fragmentation, and dynamic exclusion was set to 20 s to minimize the acquisition of fragment spectra representing already acquired precursors.

MS data processing

MS raw data were processed using MaxQuant v1.6.17.0 with default parameters. Acquired spectra were searched against the human 'one sequence per gene' database (Taxonomy ID 9606) downloaded from UniProt (12 March 2020; 20,531 sequences), and a collection of 244 common contaminants ('contaminants.fasta' provided with MaxQuant v1.6.17.0) using the Andromeda search engine integrated into MaxQuant v1.6.17.0 (ref. [28,29]). Identifications were filtered to obtain false discovery rates below 1% for both peptide spectrum matches (minimum length of seven amino acids) and proteins using a target–decoy strategy[30]. Protein quantification and data normalization relied on the MaxLFQ algorithm implemented in MaxQuant v1.6.17.0 (ref. [31]). The MS proteomics data have been deposited to the ProteomeXchange Consortium[32] via the PRIDE partner repository[33] with the dataset identifiers PXD032721 (Fig. 1a,b), PXD032740 (Extended Data Fig. 8b), PXD032741 (Fig. 4d and Extended Data Fig. 8c–e), PXD032743 (Fig. 4e), PXD032750 (Extended Data Fig. 9h–k), PXD039186 (Extended Data Fig. 10m,n), PXD039187 (Fig. 5c) and PXD039188 (Extended Data Fig. 10p,q). For protein assignment, spectra were correlated with the UniProt human database v2019 including a list of common contaminants. Searches were performed with tryptic specifications and default settings for mass tolerances in MS and MS/MS spectra. Carbamidomethyl cysteine, methionine oxidation and N-terminal acetylation were defined as fixed modifications. The match-between-run feature was used with a time window of 1 min. For further analysis, Perseus v2.0.7.0 was used and first filtered for contaminants and reverse entries as well as proteins that were only identified by a modified peptide.

Structural modelling of ubiquitinated FAM134 proteins

The previously built molecular model of FAM134B RHD[4] was extended to include an additional ten residues at the C terminus of the RHD (residues 261–270). Isopeptide bonds between the lysine (K160 and K264) and the

terminal glycine of Ub (G76) were modelled by modifying the side-chain lysine bead (SC2/+1) into the neutral backbone bead (BB/0) and restraining its distance to the terminal bead of Ub to 0.35 nm with a force constant k = 1,250 kJ mol^{-1}. Two mono-ubiquitinated and one bi-mono-ubiquitinated RHD structures were modelled (K160–Ub, K264–Ub and (K160 + K264)-Ub, respectively).

MD simulations and analysis

Coarse-grained (CG) MD simulations were prepared using the MARTINI model (v2.2)[34,35]. Initial CG structures were built using martinize.py[36]. Assignments from DSSP (Dictionary of Secondary Structures in Proteins) program were used to generate backbone restraints to preserve local secondary structure[37,38]. CG models were embedded into POPC (16:0–18:1 phosphatidylcholine (PC)) bilayers spanning the periodic simulation box in the x–y plane. Initial configurations for each system were assembled and then solvated with CG water containing 150 mM NaCl using insane.py[36]. Each system was energy minimized and equilibrated using the Berendsen thermostat[39] and barostat[40] along with position restraints on protein backbone beads, followed by production runs with a 20-fs time step. System temperature and pressure during the production phase were maintained at 310 K (unless otherwise stated) and 1 atm with the velocity rescaling thermostat[41] and the semi-isotropic Parrinello–Rahman[42] barostat, respectively. All simulations were performed using gromacs (v2019.3)[43,44]. Long-lived, highly populated RHD conformations were obtained after clustering evenly sampled conformations (n = 10,000) from each trajectory. Clusters were obtained using backbone root-mean-square deviation (cut-off = 0.8 nm) by using the gromos method[45], as implemented in the gmx_cluster tool.

Curvature induction by ubiquitinated RHDs: bicelle-to-vesicle transitions and kinetics

Discontinuous bicelle systems containing saturated DMPC (14:0 PC) and DHPC (7:0 PC) lipids were assembled as previously described[4,5]. The equilibrated ubiquitinated and non-ubiquitinated RHD molecules obtained

from simulations in the POPC bilayers (after 5 μs) were then embedded in the bicelle and solvated. One hundred replicates for each system were simulated with different initial velocities at 300 K to obtain statistics on the transition times to vesicles. Shape transformations from flat bicelles ($H = 0$ nm^{-1}) to curved vesicles ($H = 0.15$ nm^{-1}) were monitored by measuring the signed membrane curvature ($H(t)$). Lipid coordinates were fitted to spherical surfaces using least squares optimization to compute membrane curvature along simulations using MemCurv[4]. Curvature away from and towards the upper/cytoplasmic leaflet are reported as positive and negative values, respectively. The statistics of waiting times (t) for the formation of vesicles (bilayer curvature, $|H| > 0.15$ nm^{-1}) for the three systems were determined from individual replicates. We modelled the kinetics of the bilayer-to-vesicle transition using a Poisson process with a lag time ($t = t' + \tau$). The time $t' = 1/k'$ describes the Poisson process with rate k'. The constant lag time τ captures the time required for vesicle closure from the curved bilayer disc. The distributions of waiting times are thus $p(t) = k' \, e^{-k'(t-\tau)}$ for $t > \tau$. We determined the rate of vesicle formation ($k = 1/(t' + \tau)$) for different systems, from fitting the cumulative distribution function for the probability density, $P(t - \tau) = ke^{-k(t-\tau)}$ corresponding to $p(t)$, to the observed waiting time distributions estimated from replicates. We used the previously computed maximum likelihood estimates of the vesiculation rate for non-ubiquitinated RHD bicelles[4] ($k_{RHD} = 0.0018$ ns^{-1}) to compute the acceleration factors for each system ($acc = k_{sys}/k_{RHD}$). Furthermore, to show the effect of temperature on the energy barrier to form closed vesicles from flat bicelles, we also simulated 20 replicates of each system at 280 K and estimated the number of successful vesicle closure events within the simulation timescale.

Curvature sensing by ubiquitinated RHDs on buckled membranes

A CG POPC bilayer was used to tile a tessellated buckled surface using LipidWrapper[46]. The buckled membrane was solvated with CG water and ions and equilibrated in a periodic box with a fixed x–y plane (57×28 nm^2) and excess membrane area (approximately 17 nm^2). This preserved the buckled shape of the bilayer, offering a range of curvature values to be sampled by embedded proteins ($H(x,y) = -0.05 \leq 0 \leq 0.05$ nm^{-1}).

Ubiquitinated and non-ubiquitinated RHDs (in separate simulations) were initially embedded in regions with small local curvature ($H(x, y) \simeq 0$). Following another equilibration phase with proteins, the systems were simulated for more than 20 μs at 310 K. The membrane profiles of the buckled surface with and without embedded proteins were analysed by using the Monge representation to compute local principal curvatures $k_1(x, y)$ and $k_2(x, y)$, the mean curvature $H(x, y)$, and the Gaussian curvature $K_G(x, y)$ as implemented in MemCurv[4].

RHD cluster formation: dimerization, tubule deformation and membrane budding

The clustering of ubiquitinated RHD molecules was simulated under periodic boundary conditions using buckled, tubular and flat membrane structures. MD simulations were initiated after embedding one K160–Ub RHD and one K264–Ub RHD into a single membrane buckle. Initially, the two proteins were placed more than 20 nm apart. The two proteins were tracked by measuring their minimum distance to identify initial contact and cluster formation. Once the initial contacts were made and dimers were formed, residue pairwise interactions were mapped across the tethered Ub moieties (K160-Ub and K264-Ub) to identify specific Ub–Ub *trans*-contacts. We then obtained the configuration of a closed membrane tubule (length of approximately 97–100 nm; diameter of approximately 12–15 nm) used in our previous work[4]. Ten ubiquitinated RHD molecules were embedded along the length of the tubule such that the individual proteins were spaced maximally away from each other. The RHD-containing tubule was equilibrated in explicit solvent (approximately 3.6×10^6 beads) under NVT (constant volume, constant temperature ensemble) and NPT (constant pressure, constant temperature ensemble) conditions along with position restraints on protein backbone beads. A production run of 10 μs was carried out, after releasing position restraints to observe curvature-mediated protein sorting and the formation of RHD clusters. We also generated initial configurations of nine (K160 + K264)–Ub–RHDs on a 3 × 3 square grid embedded in POPC bilayers such that each protein was separated by its nearest neighbour by 10 nm. We changed the lipid-number bilayer asymmetry from $\Delta N = 0$ to $\Delta N = 300$ using the insane.py script. Following a

previously implemented method[47], we scaled the protein−protein LJ pair interaction well depth, ${{\epsilon }}_{\alpha }={{\epsilon }}_{0}+\alpha ({{\epsilon }}_{{\rm{original}}}-\,{{\epsilon }}_{0})$. A value of $\alpha = 0.65$ corresponds to a reduction in PPI strength in the MARTINI model and a value of $\alpha = 1.0$ recovers the full interaction in the MARTINI forcefield, $\epsilon_1 = \epsilon_{\text{original}}$. The resulting $2 \times 2 = 4$ bilayer configurations were then solvated with CG water containing 150 mM NaCl. We embedded the ubiquitinated and non-ubiquitinated RHD proteins in the asymmetric membranes ($30 \times 30 \times 20\,\text{nm}^3$) in a square grid and energy minimized the systems using a soft-core potential and steepest-decent algorithm to remove steric clashes with lipids. Production runs for 5 µs at NPT conditions were carried out for each simulation condition, in which position restraints were released to observe curvature-mediated protein sorting effects and Ub–Ub interactions, leading to the formation of RHD clusters. We monitored the box dimensions over time, which acted as proxies to indicate spontaneous budding. In addition, we tracked the *z*-coordinates of the centre-of-mass (COM) of all nine proteins along with the lowest and the highest points of the PO4 beads comprising the POPC bilayer. We also computed the distance matrix specifying all inter-RHD distances computed from COM positions and used hierarchical clustering with single linkage and a cut-off of 10 nm to obtain the distinct protein clusters in each frame of the simulation. We quantified the number of clusters and their sizes to identify the largest cluster for each simulation frame.

Antibody–oligonucleotide conjugation for DNA-PAINT

For exchange DNA-PAINT experiments, donkey anti-rabbit antibodies (#711-005-152, AffiniPure) and goat anti-mouse antibodies (#115-005-003, AffiniPure) were covalently labelled with the short DNA-docking strands anti-R2 (5′-ACCACCACCACCACCACCA-3′) and anti-R1 (5′-TCCTCCTCCTCCTCCTCCT-3′), respectively, using DBCO-sulfo-NHS ester chemistry[48]. In brief, concentrated secondary antibodies were incubated with a 20-fold molar excess of DBCO-sulfo-NHS ester (Jena Bioscience). After removing excess reagent, azide-functionalized DNA-docking strands were added at a tenfold molar excess and incubated overnight at 4 °C. Unbound DNA was removed from the samples concomitant with storage buffer exchange (PBS) using Amicon centrifugal

filters (100 kDa cut-off). Antibody–DNA complexes were concentrated to 5 mg ml^{-1} and stored at 4 °C.

Exchange DNA-PAINT sample preparation

For the exchange DNA-PAINT experiments, U2OS cells were seeded in Ibidi μ-slide VI chambers at 70% confluency. The cells were fixed for 30 min with pre-warmed (37 °C) 4% methanol-free formaldehyde (Sigma-Aldrich) in PBS followed by three washes with PBS. Fixed cells were then incubated in permeabilization/blocking buffer (10% FBS and 0.1% saponin) containing primary antibodies for 60 min at room temperature: rabbit anti-FAM134B (Genscript), mouse anti-LC3B or mouse anti-REEP5, each diluted 1:200. Excess primary antibodies were removed from the chambers by three washes with PBS. Cells were then incubated with custom DNA-labelled secondary antibodies in the permeabilization/blocking buffer for 60 min, followed by the removal of free antibodies by three washes with PBS. Finally, samples were post-fixed with 4% methanol-free formaldehyde in PBS for 10 min at room temperature, followed by three washes with PBS. For the exchange DNA-PAINT experiments, 125-nm gold beads (Nanopartz) were used as fiducial markers. The gold beads were sonicated for 10 min, diluted 1:30 in PBS and sonicated again for 10 min. We added 100 μl of the gold bead solution to each Ibidi μ-slide VI chamber, settled for 10 min and washed three times with PBS. The cell chambers were finally connected to a microfluidic device (Bruker). Before imaging, ATTO-655-labelled imager strands R1 or R2 (50 pM, 0.5 M NaCl and PBS, pH 8.3) were injected into the flow chamber at a flow rate of 200 μl per minute. In the exchange DNA-PAINT experiments, imager strands were exchanged between sequential imaging cycles by washing the samples with PBS and injecting new imager strands under equal conditions.

DNA-PAINT microscopy setup and data acquisition

Exchange DNA-PAINT data were captured as previously described[49] using the N-STORM super-resolution microscopy system (Nikon) equipped with a ×100 oil immersion objective (Apo, NA 1.49) and an EMCCD camera (Andor Technology). ATTO-655-conjugated oligonucleotides were excited with a collimated 647-nm laser beam (at an intensity of 1.1 kW cm^{-2}

measured at the objective) in highly inclined and laminated optical sheet mode. We acquired 20,000 consecutive frames at 10 Hz in active frame transfer mode with an EMCCD gain of 200, a pre-amp gain of 1 and at an effective pixel size of 158 nm. For astigmatism-based 3D exchange DNA-PAINT experiments, a customized cylindrical lens (RCX-39.0.38.0-5000.0-C-425-675, 10-m focal length, CVI Laser Optics) was inserted into the emission light path. NIS Elements (Nikon), LCControl (Agilent) and Micro-Manager (v1.4.22)[50] were used for optical setup control and data acquisition. FAM134B, LC3B-II and REEP5 were imaged sequentially following the microfluidic-assisted exchange of R2 and R1 imager strands.

DNA-PAINT image processing

Single-molecule localization and image reconstruction were carried out using Picasso (v0.2.8)[10]. Single-molecule localization was achieved by integrated Gaussian maximum likelihood estimation with the following parameters: minimum net gradient = 40,000, baseline = 205, sensitivity = 4.78 and quantum efficiency = 0.95. Fiducial gold bead markers were used for post-imaging drift correction and alignment of FAM134B, LC3B-II and REEP5 channels. Single-molecule point-spread functions were filtered based on the ATTO-655 single-molecule footprint (point-spread function (PSF) symmetry $0.7 < \text{full width at half maximum (FWHM)}(x)/\text{FWHM}(y) < 1.4$, intensity threshold and localization precision < 50 nm). Signals from the same origin were linked within a radius of five times the nearest neighbour based analysis (NeNa[51]) localization precision, and with a maximum dark time of eight consecutive frames. Signals arising from the fiducial marker that passed the filtering process were removed by excluding traces from the same origin with a length exceeding 20 consecutive frames. For exchange 3D DNA-PAINT experiments, a 3D calibration curve was recorded using *z*-stacks (step size = 50 nm) of surface-immobilized 0.1-µm TetraSpeck Microspheres (Thermo Fisher Scientific). The axial section was limited to 800–1,000 nm, depending on the robustness of the fit. Multi-channel single-molecule 3D localizations from exchange DNA-PAINT experiments were aligned in Picasso and visualized in ViSP (v1.0)[52]. The 3D movie (Supplementary Video 7) was generated using ViSP v1.0 open-source software[53].

Clusters and statistical analysis

FAM134B and LC3B-II nanoclusters were identified in DNA-PAINT images using the density-based spatial clustering and application with noise (DBSCAN) algorithm with a radius of 31 nm and a minimum density of ten localizations. Cluster diameters were calculated for each condition from cluster areas. Microscale FAM134B clusters were segmented from nanoscale ER-phagy initiation sites using SR-Tesseler (v1.0.0.1)[48]. Voronoi tessellation was computed by calculating the first rank order local density map from single-molecule localizations. Thresholds for cluster segmentation were determined previously for each cell using Picasso (density factor $2 > \delta < 6$, localizations $n = 800$ and minimum diameter = 100 nm). HA–FAM134B nanocluster diameters and microcluster areas were tested for normal distribution using a Shapiro–Wilk normality test, and statistical significance was determined using a non-parametric Mann–Whitney *U*-test. The relative frequency distribution of nanocluster diameters was fitted with a log-normal distribution. The mode of the log-normal distribution is a measure for the cluster diameter (*d*) and was calculated as $d={e}^{({\mu }-{\sigma }^{2})}$. OriginPro 2020 v9.7 (Origin Lab) was used for statistical analysis.

Quantitative analysis of FAM134B copy numbers in nanoscale clusters

For the quantification of FAM134B copy numbers in nanoscale clusters, the mean dark time of binding events was analysed from intensity–time traces as previously described[13]. In brief, FAM134B clusters that colocalized with LC3B-II were selected manually from DNA-PAINT images and the mean dark time of binding events was determined from the plot of cumulative histograms using Picasso built-in functions. As a calibration, we used extracellular primary–secondary antibody complexes adhering to the sample coating. The inverse of dark times, also known as the qPAINT index[54], is proportional to the number of docking strands in clusters and was used for protein copy number determination. We fitted the relative frequency distribution of inverse dark times with a log-normal distribution and determined the mode of distribution.

Statistical analysis

All experiments were performed in at least three independent experiments if not indicated otherwise. Data are presented as mean ± standard error of mean if not indicated otherwise. For statistical analysis, raw data were analysed for normal distribution with the Kolmogorov–Smirnov test or with graphical analysis using the Q–Q plot. If appropriate, we either used one-way ANOVA (with Bonferroni post-hoc test if not indicated otherwise), repeated-measures two-way ANOVA, Kruskal–Wallis *H*-test, Student's *t*-test (one-tailed or two-tailed) or the Mann–Whitney *U*-test. $P < 0.05$ were considered significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The proteomics data are deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers: ubiquitination promotes FAM134B-mediated ER-phagy (PXD032721), FAM134B homodimer ubiquitination (PXD032740), FAM134B oligomer ubiquitination and ER-phagy (PXD032741), binding partners of FAM134B WT and 17KR oligomers (PXD032743), E3 ligase regulates FAM134B ubiquitination (PXD032750), in vitro FAM134B ubiquitination (PXD039186), in vivo FAM134B ubiquitination in *AMFR*-knockdown cells (PXD039187) and in vitro FAM134B RHD and Ub–RHD–Ub ubiquitination (PXD039188); MD simulation trajectory files and corresponding parameter files are large and span long microsecond timescales and multiple replicates can only be shared upon specific requests. All the data analysis of this study is in the Supplementary information. Source data for gels and blots are provided as Supplementary information. Source data are provided with this paper.

Code availability

The analysis codes for bicelle-to-vesicle simulations and curvature computations on simulations with buckled membranes are provided in the GitHub repository (https://github.com/bio-phys/MemCurv).

References

1. Gubas, A. & Dikic, I. ER remodeling via ER-phagy. *Mol. Cell* **82**, 1492–1500 (2022).

2. Chino, H. & Mizushima, N. ER-phagy: quality and quantity control of the endoplasmic reticulum by autophagy. *Cold Spring Harb. Perspect. Biol.* **15**, a041256 (2023).

3. Khaminets, A. et al. Regulation of endoplasmic reticulum turnover by selective autophagy. *Nature* **522**, 354–358 (2015).

4. Bhaskara, R. M. et al. Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy. *Nat. Commun.* **10**, 2370 (2019).

5. Reggio, A. et al. Role of FAM134 paralogues in endoplasmic reticulum remodeling, ER-phagy, and collagen quality control. *EMBO Rep.* **22**, e52289 (2021).

6. Siggel, M., Bhaskara, R. M., Moesser, M. K., Dikic, I. & Hummer, G. FAM134B-RHD protein clustering drives spontaneous budding of asymmetric membranes. *J. Phys. Chem. Lett.* **12**, 1926–1931 (2021).

7. Jiang, X. et al. FAM134B oligomerization drives endoplasmic reticulum membrane scission for ER-phagy. *EMBO J.* **39**, e102608 (2020).

8. Chino, H., Hatta, T., Natsume, T. & Mizushima, N. Intrinsically disordered protein TEX264 mediates ER-phagy. *Mol. Cell* **74**, 909–921.e906 (2019).

9. Sauer, M. & Heilemann, M. Single-molecule localization microscopy in eukaryotes. *Chem. Rev.* **117**, 7478–7509 (2017).

10. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

11. Levet, F. et al. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat. Methods* **12**, 1065–1071 (2015).

12. Ester, M., Kriegel, H.P., Sander, J., Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (AAAI Press, 1996).

13. Jungmann, R. et al. Quantitative super-resolution imaging with qPAINT. *Nat. Methods* **13**, 439–442 (2016).

14. Croucher, D. R. et al. Bimolecular complementation affinity purification (BiCAP) reveals dimer-specific protein interactions for ERBB2 dimers. *Sci. Signal.* **9**, ra69 (2016).

15. Grumati, P. et al. Full length RTN3 regulates turnover of tubular endoplasmic reticulum via selective autophagy. *eLife* **6**, e25555 (2017).

16. Gubas, A. & Dikic, I. A guide to the regulation of selective autophagy receptors. *FEBS J.* **289**, 75–89 (2022).

17. Chen, Z., Du, S. & Fang, S. gp78: A multifaceted ubiquitin ligase that integrates a unique protein degradation pathway from the endoplasmic reticulum. *Curr. Protein Pept. Sci.* **13**, 414–424 (2012).

18. Li, W. et al. Mechanistic insights into active site-associated polyubiquitination by the ubiquitin-conjugating enzyme Ube2g2. *Proc. Natl Acad. Sci. USA* **106**, 3722–3727 (2009).

19. Foronda, H. et al. Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy. *Nature* https://doi.org/10.1038/s41586-023-06090-9 (2023).

20. Liang, J. R. et al. A genome-wide ER-phagy screen highlights key roles of mitochondrial metabolism and ER-resident UFMylation. *Cell* **180**, 1160–1177.e20 (2020).

21. Hubner, C. A. & Dikic, I. ER-phagy and human diseases. *Cell Death Differ.* **27**, 833–842 (2020).

22. Gordon, W. R. et al. Effects of S1 cleavage on the structure, surface export, and signaling activity of human Notch1 and Notch2. *PLoS ONE* **4**, e6613 (2009).

23. Izadi, M. et al. Cobl-like promotes actin filament formation and dendritic branching using only a single WH2 domain. *J. Cell Biol.* **217**, 211–230 (2018).

24. Chaudhary, S., Pak, J. E., Gruswitz, F., Sharma, V. & Stroud, R. M. Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. *Nat. Protoc.* **7**, 453–466 (2012).

25. Schneider, K. et al. ProSAP1 and membrane nanodomain-associated syndapin I promote postsynapse formation and function. *J. Cell Biol.* **205**, 197–215 (2014).

26. Seemann, E. et al. Deciphering caveolar functions by syndapin III KO-mediated impairment of caveolar invagination. *eLife* **6**, e29854 (2017).

27. Wolf, D. et al. Ankyrin repeat-containing N-Ank proteins shape cellular membranes. *Nat. Cell Biol.* **21**, 1191–1205 (2019).

28. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

29. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

30. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).

31. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

32. Deutsch, E. W. et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).

33. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

34. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

35. Monticelli, L. et al. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

36. Wassenaar, T. A., Ingolfsson, H. I., Bockmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

37. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

38. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

39. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).

40. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).

41. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

42. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

43. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).

44. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).

45. Daura, X. et al. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236–240 (1999).

46. Durrant, J. D. & Amaro, R. E. LipidWrapper: an algorithm for generating large-scale membrane models of arbitrary geometry. *PLoS Comput. Biol.* **10**, e1003720 (2014).

47. Benayad, Z., von Bulow, S., Stelzl, L. S. & Hummer, G. Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput.* **17**, 525–537 (2021).

48. Schroder, M. S. et al. Imaging the fibroblast growth factor receptor network on the plasma membrane with DNA-assisted single-molecule super-resolution microscopy. *Methods* **193**, 38–45 (2021).

49. Jungmann, R. et al. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–318 (2014).

50. Edelstein, A. D. et al. Advanced methods of microscope control using muManager software. *J. Biol. Methods* **1**, e10 (2014).

51. Endesfelder, U., Malkusch, S., Fricke, F. & Heilemann, M. A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* **141**, 629–638 (2014).

52. Glogger, M., Spahn, C., Enderlein, J. & Heilemann, M. Multi-color, bleaching-resistant super-resolution optical fluctuation imaging with oligonucleotide-based exchangeable fluorophores. *Angew. Chem. Int. Ed.***60**, 6310–6313 (2021).

53. El Beheiry, M. & Dahan, M. ViSP: representing single-particle localizations in three dimensions. *Nat. Methods* **10**, 689–690 (2013).

54. Jayasinghe, I. et al. True molecular scale visualization of variable clustering properties of ryanodine receptors. *Cell Rep.* **22**, 557–567 (2018).

Acknowledgements

We thank A.-C. Jacomin, A. Gubas and D. Höller for critical reading of the manuscript; all the members of the I.D. laboratory for their support and constructive discussion; the q-Proteomics platform at the Institute of Biochemistry II (Frankfurt), G. Tascher, T. Colby and I. Matic for the proteomics support; and A. Kreusch, M. Öhler, S. Berr and K. Gluth in the B.Q. and M.M.K. laboratories for their practical help with cloning, protein purification, liposome generation and liposome shaping assays. We thank the Central Electron Microscopy facility of the Max Planck Institute of Biophysics for technical support. G.H. thanks the Max Planck Society for support. R.M.B. and G.H. thank the Center for Supercomputing, GUF, and the MPCDF, Garching for computing resources. This research is supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), project numbers 259130777-SFB 1177 to I.D., V.D., M.H. and G.H., and QU116/9-1 to B.Q., KE685/7-1 to M.M.K. and HU800/15-1 to C.A.H., the European Union (ERC, ER-REMODEL, 101055213), the HMWK-funded cluster grant ENABLE, the Leistungszentrum Innovative Therapeutics (TheraNova) funded by the Fraunhofer Society and the Hessian Ministry of Science and Art, and grants from Else Kroener

Fresenius Stiftung and Dr. Rolf M. SchwieteStiftung to I.D. S.K.K. is supported by a postdoctoral fellowship from EMBO (ALTF 199-2021).

Author information

Author notes

1. These authors contributed equally: Alexis González, Adriana Covarrubias-Pinto

Authors and Affiliations

1. Institute of Biochemistry II, Faculty of Medicine, Goethe University Frankfurt, Frankfurt am Main, Germany

   Alexis González, Adriana Covarrubias-Pinto, Ramachandra M. Bhaskara, Santosh K. Kuncha, Audrey Xavier, Mohit Misra, Marina E. Hoffmann & Ivan Dikić

2. Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

   Ramachandra M. Bhaskara, Santosh K. Kuncha, Audrey Xavier, Mohit Misra & Ivan Dikić

3. Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt am Main, Germany

   Ramachandra M. Bhaskara & Gerhard Hummer

4. Institute of Physical and Theoretical Chemistry, Goethe University Frankfurt, Frankfurt, Germany

   Marius Glogger, Ashwin Balakrishnan & Mike Heilemann

5. Institute of Biochemistry I, Jena University Hospital, Friedrich Schiller University Jena, Jena, Germany

Eric Seemann, Britta Qualmann & Michael M. Kessels

6. Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany

   Bastian Bräuning & Brenda A. Schulman

7. Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, Frankfurt, Germany

   Volker Dötsch

8. Institute of Human Genetics, University Hospital Jena, Friedrich Schiller University, Jena, Germany

   Christian A. Hübner

9. Institute of Biophysics, Goethe University Frankfurt, Frankfurt am Main, Germany

   Gerhard Hummer

10. Fraunhofer Institute of Translational Medicine and Pharmacology, Frankfurt am Main, Germany

    Ivan Dikić

Contributions

The present study was conceived by I.D., A.G. and A.C.-P. The experiments and analysis were performed by A.G., A.C.-P., R.M.B., M.G., S.K.K., A.X. and E.S. A.G., A.C.-P. and E.S. prepared all the constructs. A.G. and A.C.-P. carried out all the cell biology and biochemical experiments, generated all the inducible cell lines and samples for MS experiments and analysed the results, including MS data. M.E.H. performed the initial analysis of MS data. R.M.B. performed structural modelling, MD simulations and analysis with the support of G.H. M.G. and A.B. carried out the super-resolution experiments and analysis (under the supervision of M.H.). Protein purification was carried out by M.M., B.B. (under the supervision

of B.A.S.), S.K.K. and E.S. Membrane-shaping assays and data analysis were carried out by S.K.K., A.X., A.C.-P., A.G., E.S., M.M.K., B.Q. and C.A.H. A.G., A.C.-P., R.M.B. and I.D. wrote the manuscript with contributions from all authors.

Corresponding author

Correspondence to Ivan Dikić.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Constitutive and inducible FAM134B-RHD ubiquitination regulates ER-phagy.

a, TUBE-2 pulldown assay showing increased ubiquitination of endogenous FAM134B following Torin 1 treatment. Cells were treated with DMSO (control), 200 nM BafA1 for 6 h, 250 nM Torin 1 for 6 h, or a combination

of 250 nM Torin 1 and 200 nM BafA1 for 6 h. Protein samples were analysed by SDS-PAGE and western blotting, as indicated. **b**, Densitometric quantification of ubiquitinated FAM134B (FAM134B-Ub) in panel **a** (Data are mean ± s.d.; $n = 3$ independent experiments, one-way ANOVA, Bonferroni post-hoc test). **c**, Cells were treated with DMSO (control), 200 nM BafA1 for 6 h or 10 µM MG132 for 6 h. Detergent-soluble extracts were analysed by western blot with antibodies against FAM134B and UbP4D1. The panels show representative immunoblots. **d**, Cycloheximide (50 µg/ml) chase for 0–6 h in HeLa cells without or with 200 nM BafA1. Detergent-soluble extracts were analysed by western blot with antibodies against FAM134B and vinculin. **e**, Densitometric quantification of FAM134B (normalised to vinculin) in panel d (Data are mean ± s.d.; $n = 3$ independent experiments, one-way ANOVA, Bonferroni post-hoc test). **f**, Cells were treated with DMSO (control) or 10 µM TAK243 for 1, 2 and 4 h before TUBE-2 pulldown assays. Endogenous ubiquitination of FAM134B was detected by western blot ($n = 1$ experiment). **g**, Cycloheximide (50 µg/ml) chase for 0–6 h in HeLa cells with or without 10 µM TAK243. Detergent-soluble extracts were analysed by western blot with antibodies against FAM134B, UbP4D1 and vinculin. **h**, Densitometric quantification of FAM134B (normalised to vinculin) in panel g (Data are mean ± s.d.; $n = 3$ independent experiments; One-way ANOVA, Tuckey's post-hoc test). **i**, Ubiquitination assay in cells showing FAM134B-RHD ubiquitination despite the replacement of eight conserved lysine residues with arginine (myc-Ub-IP). LC3B-II binding to FAM134B and the formation of high-molecular-weight species (oligomers, SDS-resistant) of FAM134B were not affected by 8KR (Flag FAM134B-IP) ($n = 1$ experiment). **j**, Co-localisation of HA-FAM134B/LC3B$^+$ puncta per cell in cells expressing FAM134B-WT or 8KR. Number of cells per condition for HA-FAM134B WT: 2039 (DMSO), 2366 (BafA1), 1515 (EBSS), 2280 (Torin1), 1987 (Torin1+BafA1). Number of cells per condition for HA-FAM134B-8KR: 2416 (DMSO), 2494 (BafA1), 1753 (EBSS), 2652 (Torin1), 2580 (Torin1+BafA1). Bars represent means ± s.d. ($n = 6$ independent experiments; One-way ANOVA, Bonferroni post-hoc test). **k**, Isolated membranes of cells expressing HA-FAM134B WT or 17KR were treated with 0.3 mM EGS and analysed by SDS-PAGE. Oligomers were visualised by western blot. ($n = 1$ experiment). **l**, U2OS cells stably expressing HA-FAM134B-WT under the control of a doxycycline-inducible promoter were

treated with DMSO (control) or 200 nM BafA1 for 6 h. **m**, HeLa cells stably expressing HA-FAM134B-WT or the LIR mutant under the control of a doxycycline inducible promoter were treated with DMSO (control) or 200 nM BafA1 for 6 h. ($n = 1$ experiment). **n**, TUBE-2 pulldown assay showing the accumulation of endogenous ubiquitinated HA-FAM134B LIR-mutant. The panels show representative immunoblots.

Source data

Extended Data Fig. 2 Ubiquitin-membrane and ubiquitin-RHD interactions.

a, Time series of the number of contacts between CG-beads of Ubiquitin and POPC bilayers in MD simulations of mono-ubiquitinated RHDs. **b**, Snapshots from MD simulations of K160-Ub (left) and K264-Ub (right) showing the structures of the Ub moieties (in cyan and blue), and their interactions with the membrane (orange beads). Modified lysines (red) show the relative orientations of the Ub-moieties with respect to the amphipathic helices (yellow) and the cytosolic portion of the RHD. **c**, Time series of the number of contacts between CG beads of RHD and covalently linked Ub moieties. **d**, Residue-wise contact maps between Ub and RHDs. The solid red lines indicate the position of the lysine residue on the RHD linked to Ub. **e**, Time series of the number of contacts between CG beads of Ubiquitin and POPC bilayers in MD simulations of bi-mono-ubiquitinated RHD, (K160+K264)-Ub. **f**, Snapshot from MD simulation of (K160 + K264)-Ub variant showing how bi-mono-ubiquitination of the RHD reduces the interactions between ubiquitin and POPC bilayer. The two Ub-moieties are involved in intra-molecular/cis interactions on top of the cytosolic face of the RHD. **g**,**h** Time series (**g**) and probability distribution (**h**) of the radii of gyration, R_g, sampled by ubiquitinated and non-ubiquitinated FAM134B-RHDs during the 10-µs MD simulations.

Extended Data Fig. 3 Ubiquitinated RHDs induce membrane curvature.

a–**d**, Time series of mean curvature (H) of bicelle containaing single FAM134B RHD. One hundred MD simulations for each system were

initiated with different initial velocities to study curvature induction (positive/negative) and transition of the bicelles into closed vesicles (H = +0.14 nm^{-1}) at 300 K. **e**, Violin plots of estimated waiting times required for bicelle-to-vesicle transitions induced by WT and Ub-variants of FAM134B-RHD. n = 92 (FAM134B-RHD), 100 (K160-Ub), 99 (K264-Ub), 100 (K160+K264-Ub) runs **f**, Comparison of mean ± SEM curvature time traces (black line ± shaded region) for each system shows that Ub-RHDs induce bicelles to curve swiftly, resulting in faster kinetics. **g**, Table showing the kinetics of the *in silico* curvature induction process. The numbers n_+ and n_- indicate bicelle transitions resulting in bilayer curvature away from and towards the upper/cytoplasmic leaflet. Observed waiting times (t) for vesicle formation were recorded as the time taken for the bilayer or bicelle discs to reach a curvature of $|H| = 0.14$ nm^{-1}. The vesicle formation rate $k_{sys} = 1/(t' + \tau)$ was estimated from exponential fits of the cumulative distribution functions of Poisson-distributed waiting times (t') and a constant lag time (τ). The acceleration in vesicle formation due to a protein inclusion was estimated as the ratio k_{sys}/k_{WT}. **h**–**k**, Time series of bicelle mean curvature (H) for 20 simulations initiated at 280 K for each system. Fewer bicelles with embedded K160-Ub and WT-RHD proteins (4/20 and 5/20) transitioned into closed vesicles at 280 K.

Extended Data Fig. 4 Curvature preference of ubiquitinated FAM134B-RHDs.

a,**b**, Selected snapshots from 20-µs MD simulations of single K160-Ub and K264-Ub variants on buckled POPC bilayers. The Ub-variants sample regions of high mean curvature and preferentially occupy the top of the buckle. **c**, Histograms of the mean curvature H(x, y) sampled by FAM134B RHD (green), K160-Ub (cyan) and K264-Ub (steel blue) in coarse-grained simulations (1-ns intervals for 20 µs) indicate a preference for highly curved regions of the buckle. For reference, the distribution of local mean curvature values on the empty buckled membrane (red) was estimated by random sampling of points in the *xy* plane, ignoring small curvature corrections. The time-averaged values of H(x, y) for each system are also provided. **d**, Confocal fluorescence microscopy analysis of cells transiently expressing HA-tagged FAM134B $RHD_{90–264}$ and the chimaera

Ub-RHD$_{90–264}$-Ub. Fixed cells were stained with anti-HA (green) and anti-REEP5 (red). Scale bar = 10µm. **e**, Co-localisation analysis of RHD/REEP5+ puncta per cell in cells expressing FAM134B RHD$_{90-264}$ or Ub-RHD$_{90-264}$-Ub (bar plots of data are presented, statistical significance was determined using a non-parametric two-tailed Mann-Whitney U-test, *n* = 10 cells per condition) Bars represent means ± s.d. **f**, Ubiquitin moieties tethered to different RHDs trigger inter-molecular Ub-Ub interactions, which induce RHD clustering and dimerisation on top of the buckle (left) after ~18 µs. (right) Time series of inter-molecular Ub-Ub contacts across the two molecules. The inset shows the average Ub-Ub contact map stabilising the dimer structure (21–25 µs). **g**, Snapshots of membrane tubule with 10 Ub-RHDs (left) Snapshots of Ub-RHD cluster formation that locally deform membrane tubules from ideal cylindrical geometry. Coloured squares (red, blue and green) highlight the number of molecules (dimers/trimers) that form the clusters locally on different parts of the tubule. Ubiquitinated RHD clusters shape the tubule (right). Zoomed RHD clusters show Ub-Ub interactions (blue moieties).

Source data

Extended Data Fig. 5 Ub-Ub interactions from MD simulations of ubiquitinated RHDs.

9 (K160+K264)-Ub-RHD molecules were simulated in a 3x3 square grid embedded in POPC bilayer under 4 different simulation conditions (*ΔN* = 0 and 300, and *α* = 1 and 0.65) to quantify the Ub-Ub interactions (schematic at the top). **a-d**, The two Ub moieties of (K160+K264)-Ub-RHD molecules are engaged in both cis/intra-molecular (left) and trans/inter-molecular interactions (right). Interactions under each simulation condition were quantified by computing an average Ub-Ub contact map, \({{\rm{UU}}}_{{\rm{cnts}}}=\left[{\sum }_{{i\in {Ub}}_{1}}{\sum }_{{j\in {Ub}}_{2}}\sigma (|{r}_{{ij}}|)\right]\), where the sums extend over residue COM positions of interacting Ub moieties and \(\sigma (|{r}_{{ij}}|)\) is a smooth sigmoidal counting function to limit interactions below the cut-off ($r_{ij} \leq 10$ Å). Interaction maps were averaged over only the bound states (\(\sum {{\rm{UU}}}_{{\rm{cnts}}}\ge 5\)) for all nine intra-molecular Ub-Ub pairs (left) and 144 inter-molecular Ub-Ub pairs (right),

respectively. **e**, Box plots showing the distribution of lifetimes of trans/inter-molecular Ub-Ub interactions from each of the four simulation conditions. The lifetime for each trans-Ub-Ub bound state was estimated as a contiguous stretch (in ns) where (($\sum {{\rm{UU}}}_{{\rm{cnts}}}\ge 5$) for 10 ns). Lifetimes from a total of 216, 276, 143 and 304 trans-Ub-Ub bound states were sampled over the course of 5-μs runs for each of the four simulation conditions, respectively.

Extended Data Fig. 6 Map of Ub-Ub interactions.

a, (Top) Schematic showing the cis/intra-molecular Ub-Ub interactions and **b**, the trans/inter-molecular Ub-Ub interactions mapped onto the 3D structure of Ub. Strongly interacting (red), moderately interacting (white), and weakly interacting (blue) sites are coloured, and the cartoon backbone size is scaled accordingly. (Bottom) The same interactions are mapped along the sequence of Ub to highlight the various secondary structural elements and residues involved, revealing that hairpins β12 and β34, along with the C-terminal region of Ub, dominate the intra-molecular interactions of K160-Ub and K264-Ub. However, trans-Ub-Ub interactions are spread throughout the Ub sequence, indicating their non-specific nature.

Extended Data Fig. 7 Effect of RHD ubiquitination on the flux of ER-phagy and FAM134B cluster size.

a, Representative confocal images of HA-FAM134B WT and 17KR (red) co-stained with LC3B (green). **b**, Co-localisation of HA-FAM134B/LC3B^{+} puncta and the **c**, corresponding area (μm^2) were lower for 17KR compared to WT. The data are representative of three independent experiments in which the total number of cells per condition were $n_{\rm cells}$ = 488 WT, 392 17KR. (Data are mean ± s.d.; two-tailed unpaired Student’s t-test). **d**, Representative negative-stain transmission electron microscopy (TEM) images of remodelled proteoliposomes (scale bars = 200 nm). Empty liposomes were incubated with purified GST empty, GST-FAM134B-WT or GST-FAM134B-17KR for 18 h at 25 °C. Images show examples of representative proteoliposomes. **e**, Violins shows the box-plots with median value, white dot, interquartile range (black shaded region), min and max values (1.5 x interquartile region) and mirrored probability density estimates

on sides. (WTmean = 28.25; 17KR mean = 27.73; GSTmean = 128.78), GST empty (n = 167), GST-FAM134B-WT (n = 277) or GST-FAM134B-17KR (n = 297); Kruskal-Wallis/Dunn's post-hoc test. **f**, Quantitative TEM analyses of diameters of freeze-fractured liposomes that were incubated with either GST control, GSTFAM134B WT or GSTFAM134B17KR. Data, mean ± SEM presented as normalized to the mean liposome diameter of GST control. Two independent liposome preparations and experiments. Kruskal-Wallis/Dunn's post-test. (n = 310 for WT, n = 250 for 17KR and n = 208 for GST control). **g**, Schematic representation of the ER-phagy reporter system RFP-GFP-KDEL. **h**, Representative confocal images of U2OS TRex stable cell lines co-expressing RFP-GFP-KDEL with either HA-FAM134B WT or HA FAM134B 17KR. Cells were treated for 16 h with 1 µg/ml doxycycline to induce the expression of both proteins. Cells were fixed, permeabilised, and stained for HA and LC3B. Bar = 10 µm. **i**, ER-phagy flux was quantified as the ratio between RFP^{+}/GFP^{-} and RFP^{+}/GFP^{+} puncta, quantified using CQ1 software. Cells were treated with DMSO (control), 200 nM BafA1 for 6 h, EBSS for 6 h, 250 nM Torin 1 for 6 h or 250 nM Torin 1 plus 200 nM BafA1 for 6 h. Data are means ± s.d. of n = 6 independent experiments, in which the number of RFP-GFP-KDEL cells per condition are: 2366 (DMSO), 2202 (BafA1), 1460 (EBSS), 2228 (Torin1), 2378 (Torin1+BafA1). Number of RFP-GFP-KDEL/HA-FAM134B WT cells: 670 (DMSO), 773 (BafA1), 578 (EBSS), 986 (Torin1), 747 (Torin1+BAfA1), Number of HA-FAM134B 17KR cells: 1160 (DMSO), 1151 (BafA1), 1313 (EBSS), 1353 (Torin1), 1295 (Torin1+BafA1). One-way ANOVA, Bonferroni post-hoc test. **j**, RFP-GFP-KDEL, RFP-GFP-KDEL/HA-FAM134B WT and RFP-GFP-KDEL/HA-FAM134B 17KR cells were left untreated or treated with EBSS for 8 h. Detergent-soluble extracts were analysed by western blot using antibodies against RFP, HA, REEP5 and β-actin. (n = 1 experiment). **k**, The E1 inhibitor decreases the flux of ER-phagy in mCherry-GFP-FAM134B-WT cells induced with Torin 1 (Data are mean ± s.d.; one-way ANOVA, Bonferroni post-hoc test). ER-phagy flux was quantified as the ratio between $mCherry^{+}/GFP^{-}$ and $mCherry^{+}/GFP^{+}$ puncta, quantified using CQ1 software. Data are means ± s.d. of n = 5 independent experiments in which the number of cells per condition were: (DMSO) 837 basal, 1072 BafA1, 1038 Torin1, 966 BafA1+Torin1. Number of cells (10 µM TAK243): 729 basal, 1174 BafA1, 1060 Torin1, 1121 Torin1+BafA1. **l**, Two-colour DNA-PAINT super-resolution image of HA-

FAM134B (magenta, R2-ATTO655) and the autophagosomal membrane marker LC3B-II (green, R1-ATTO655) (i). White box indicates the magnified region shown in (ii). (iii) Point localisations of HA-FAM134B from the magnified region shown in (ii) and corresponding Voronoi diagrams (blue polygons) with red line representing FAM134B cluster contour (iv). Clusters are identified based on previously determined thresholds (density factor, minimum number of localisations and minimal distance parameter). Scale bars = 10 µm (i) and 1 µm (ii–iv). **m**, Box plot of HA-FAM134B-WT and HA-FAM134B-17KR nanoscale and microscale cluster sizes. For nanoscale clusters, ubiquitination-deficient FAM134B significantly reduces the cluster diameter (35–202 nm, median = 85 nm) compared to its WT counterpart (33–286 nm, median = 114 nm). Nanoscale cluster ($n_{\text{cells}} = 4$, $n_{\text{WTclusters}} = 1278$; $n_{\text{17KRclusters}} = 1255$). For larger microscale clusters, significantly larger areas were detected for HA-FAM134B-WT (0.017–0.20 µm^2, median = 0.08 µm^2) compared to the 17KR mutant (0.01–0.21 µm^2, median = 0.03 µm^2). Torin 1 treatment further increased the HA-FAM134B cluster area with the effect being stronger for ubiquitinated HA-FAM134B-RHD ($\text{median}_{\text{WT}} = 0.23$ µm^2, $\text{median}_{\text{17KR}} = 0.16$ µm^2). Quantitative analysis of nanoscale clusters was carried out using the DBSCAN algorithm and microscale clusters were identified using SR-tessellation. Box-plots of FAM134B wildtype (magenta, grey dots) and FAM134B 17 KR (green, grey squares) nanoscale cluster diameters (left panel) and microscale cluster areas (right panel, grey dots) showing median values (horizontal lines in boxes), the interquartile ranges (width of the boxes) and whiskers defining minimum and maximum values (excluding outliers). A non-parametric one-tailed Mann-Whitney U-test was applied to the data. **n**, HA-FAM134B forms nanoscale clusters within the ER network. Two-colour super-resolution image of HA-FAM134B-WT (magenta, R2-ATTO655) and ER-membrane marker REEP5 (green, R1-ATTO655), with (ii) showing the magnified region from box (i). Scale bars = 10 µm (i) and 1 µm (ii).

Source data

Extended Data Fig. 8 Protein interactors of FAM134B/FAM134C-containing oligomers.

a, HEK 293T cells were transfected with a control plasmid (GFP), V1-FAM134B-WT, V2-FAM134B-WT, V1-FAM134C-WT or V2-FAM134C-WT. Cells were also co-transfected with V1-FAM134B-WT and V2-FAM134B-WT or V1-FAM134C-WT and V2-FAM134C-WT. The GFP trap was analysed by western blot. ($n = 1$ experiment). **b**, Ubiquitinated lysine residues identified by proteomics analysis in the RHD of immuno-isolated dimeric FAM134B: diGly peptides were significantly enriched ($\log_2$ enrichment > 2.0 and $-\log_{10}$ p value > 1.3, one-tailed unpaired Student's test). $n = 3$ independent experiments. Schematic of the FAM134B-RHD showing the localisation of ubiquitinated lysine residues. **c**, Single-sided volcano plot of the quantitative label-free interactome of FAM134C homodimers and **d**, FAM134B/FAM134C heterodimers depicting identified RHD-containing ER proteins (blue), autophagy-related proteins (green), and components of ubiquitination machinery (red) ($\log_2$ enrichment > 2.0 and $-\log_{10}$ p value > 1.3). Data are means ± s.d. of $n = 3$ independent experiments. **e**, Heat map comparing the interaction of RHD-containing ER proteins, autophagy-related proteins and the ubiquitination machinery with WT FAM134B homodimers, FAM134C homodimers and FAM134B/FAM134C heterodimers (immuno-isolated using BiCAP). Interaction partners with $\log_2$ enrichment > 2.0 and $-\log_{10}$ p value > 1.3 were plotted. $n = 3$ independent experiments, one-tailed unpaired Student's test. **f**, Venn diagram of interactors of FAM134B homodimers, FAM134C homodimers and FAM134B /FAM134C heterodimers. Numbers represent significantly enriched interaction partners ($\log_2$ enrichment > 2.0 and $-\log_{10}$ p > 1.3, one-tailed unpaired Student's test). $n = 3$ independent experiments. **g**, Annotation enrichment analysis of the interactome of FAM134B and FAM134C heterodimers. Bars represent significantly enriched gene ontology biological process (GOBP), gene ontology cellular component (GOCC), gene ontology molecular function (GOMF), and domain enrichment (Pfam). **h**, Confocal fluorescence microscopy analysis of the BiFC signal produced by interactions between V1-FAM134B and V2-FAM134B. Fixed cells expressing V1-FAM134B and V2-FAM134B were stained for LC3B (red) and p62 (blue). Scale bar = 10 µm. **i**, Pearson's correlation coefficients obtained from the co-localisation analysis of fluorescent signals representing FAM134B clusters and p62 or LC3B. Data are means ± s.d. of $n = 10$ cells per analysis.

Source data

Extended Data Fig. 9 Analysis of the functional interaction between AMFR and FAM134B in cells.

a, Confocal fluorescence microscopy analysis of the BiFC signals following the interaction between V1-FAM134B and V2-FAM134B. Fixed cells expressing V1-FAM134B and V2-FAM134B were stained for AMFR (red) and LC3B (blue). Scale bar = 10 µm. **b**, The fluorescence intensity distribution reveals that FAM134B clusters (V1FAM134B+V2FAM134B) co-localise with endogenous AMFR and LC3B in punctate structures. **c**, Ubiquitination assay of HA-FAM134B in cells co-expressing WT-AMFR-Flag or the catalytically inactive AMFR-Flag (C356G H361A) variant. **d**, Densitometric quantification of western blot signals for ubiquitinated HA-FAM134B following co-expression with WT-AMFR-Flag or the catalytically inactive AMFR-Flag (C356G H361A) variant as presented in Fig. 9c. Data are means ± s.d. of $n = 3$ independent experiments, two-tailed unpaired Student's t test. **e**, Schematic of the BiCAP method to study the functional interaction between FAM134B and AMFR. Full-length FAM134B was fused to the C-terminal of the non-fluorescent N-terminal (V1) fragment of Venus protein (V1-FAM134B), whereas full-length AMFR was fused to the N-terminal of the non-fluorescent C-terminal (V2) fragment (AMFR-V2). **f**, Confocal microscopy analysis of fixed cells co-expressing V1-FAM134BWT and AMFR-V2 WT or V1-FAM134BWT and AMFR-V2 C356G H361A stained for Ub and LC3B (magenta). Scale bar = 10 µm. **g**, Pearson's correlation coefficients obtained from co-localisation of the fluorescent signals representing refolded Venus and Ub(FK2) in cells co-expressing V1-FAM134BWT and AMFR-V2 WT or V1-FAM134BWT and AMFR-V2 C356G H361A (Extended Data Fig. 9e), (Two-tailed non-parametric Mann-Whitney U-test was applied to the data) Data are means ± s.d. $n_{\text{V1FAM134B/V2AMFRWT}} = 68$ puncta and $n_{\text{V1FAM134B/V2AMFR C356G H361A}} = 52$ puncta. **h**, Single-sided volcano plot of the quantitative label-free interactome of the affinity-purified (using BiCAP) WT-AMFR-V2/V1-FAM134BWT complex. RHD-containing ER proteins (blue), autophagy-related proteins (green), and ubiquitination machinery (red) ($\log_2$ enrichment > 2.0 and $-\log_{10}$ value > 1.3. Data are means of $n = 3$ independent experiments; one-tailed unpaired Student's test. **i**, Venn diagram of

interactors of FAM134B homodimers and AMFR/FAM134B heterodimers. Numbers represent significantly enriched peptides ($\log_2$ enrichment > 2.0 and $-\log_{10}$ p value > 1.3. Data are means of $n = 3$ independent experiments, one-tailed unpaired Student's test. **j**, Comparison of Torin 1-induced ubiquitination of FAM134B within complexes with either AMFR-WT (WT-AMFR-V2/V1-FAM134B) or the catalytically inactive AMFR mutant (C356G H361A AMFR-V2/V1-FAM134B). Graphs show the abundance of FAM134B peptides carrying a diGly modification expressed as intensities. Data are means ± s.d. of $n = 3$ independent experiments and the identified diGly peptides intensities were normalised to the total intensities of modified and non-modified peptides (two-tailed unpaired Student's test). **k**, Heat map comparing the interaction of RHD-containing ER proteins, autophagy-related proteins and the ubiquitination machinery of WT-AMFR-V2/V1-FAM134B and AMFR C356G H361A-V2/V1-FAM134B complexes. Interaction partners with $\log_2$ enrichment > 2.0 and $-\log_{10}$ p value > 1.3 were plotted. Data are means of $n = 3$ independent experiments, one-tailed unpaired Student's test.

Source data

Extended Data Fig. 10 Analysis of the functional interaction between AMFR and FAM134B in cells and *in vitro*.

a, HeLa cells were transfected with control siRNA (siNT), siRNA#1 or siRNA#2 targeting AMFR (siAMFR) for 72 h. Detergent-soluble protein extracts were analysed by western blot using antibodies against FAM134B, AMFR or vinculin (loading control). **b**, Densitometric quantification of endogenous FAM134B (normalised to vinculin) in panel a (Data are means ± s.d. of $n = 5$ independent experiments, One-way ANOVA, Bonferroni post-hoc test). **c**, Confocal fluorescence microscopy of U2OS cells stably expressing HA-FAM134BWT transfected with either control siRNA (siNT) or siRNA#1 targeting AMFR (siAMFR) for 72 h, followed by incubation with 250 nM Torin 1 for 6 h. Cells were fixed and stained for HA-FAM134B and endogenous LC3B, respectively. **d**, Quantification of HA-FAM134B-WT/LC3B-II-containing puncta per cell of images in panel c. Scatter plot graphs represent means ± s.d. ($n_{siNT} = 24$ cells, $n_{siAMFR} = 33$ cells; two-tailed Mann-Whitney-U-test). **e**, HeLa cells were treated with 250 nM Torin 1 for

the indicated time (h). Densitometric quantification of endogenous AMFR (normalised to vinculin), as presented in Fig. 5a. Data are means ± s.d. of *n* = 3 independent experiments; two-way ANOVA, Bonferroni post-hoc test. **f**, U2OS cells stably expressing HA-FAM134BWT or HA-FAM134B-17KR were incubated with 250 nM Torin 1 for 0, 2, 4, 6 and 8 h. Detergent-soluble extracts were analysed by western blot using antibodies against HA, AMFR and vinculin. **g**,**h**, Densitometric quantification of HA-FAM134B and AMFR (normalised to vinculin) from panel f. Data are means ± s.d. of *n* = 3 independent experiments, two-way ANOVA, Bonferroni post-hoc test. **i**, U2OS cells stably expressing HA-FAM134BWT or HA-FAM134B-LIR incubated with 250 nM Torin 1 for 0, 2, 4, 6 and 8 h. Detergent-soluble extracts were analysed by western blot using antibodies against HA, AMFR and vinculin. **j**, Densitometric quantification of AMFR (normalised to vinculin) in panel i. Data are means ± s.d. of *n* = 3 independent experiments, two-way ANOVA, Bonferroni post-hoc test. **k**, and **l**, Purification of AMFR from HEK 293T cells. **m**,**n**, Mass spectrometry (MS) analysis of the *in vitro* ubiquitination of full length GST-tagged FAM134B using recombinant AMFR. (Data are means ± s.d. of *n* = 3 independent experiments; two-tailed unpaired Student's t-test). **o**–**q**, MS analysis of the *in vitro* ubiquitination of His-$RHD_{90–264}$-Strept-II and His-Ub-$RHD_{90–264}$-Ub-Strept-II using recombinant AMFR. Data are means ± s.d. of *n* = 3 independent experiments; two-tailed unpaired Student's t-test. **r**, Immunodetection of native His-$RHD_{90–264}$-Strept-II following ubiquitination by AMFR. The ubiquitination reaction was analysed by western blot after blue native polyacrylamide gel electrophoresis (BN-PAGE) using antibodies against His_6 or UbP4D1. Control reaction (in the presence of AMFR, no ATP).

Source data

Supplementary information

Supplementary Information

This file contains immunoblot source data. Uncropped immunoblots of the main Figures (1a-1c) and extended data (1d-1t).

Reporting Summary

Peer Review File

Supplementary Video 1

Representative coarse-grained MD trajectory of bicelle (DMPC + DHPC lipids)-to-vesicle transition induced by embedded K160-Ub FAM134B-RHD.

Supplementary Video 2

Representative coarse-grained MD trajectory of bicelle (DMPC + DHPC lipids)-to-vesicle transition induced by embedded K264-Ub FAM134B-RHD.

Supplementary Video 3

Representative coarse-grained MD trajectory of bicelle (DMPC + DHPC lipids)-to-vesicle transition induced by embedded (K160+K264)-Ub FAM134B-RHD.

Supplementary Video 4

Coarse-grained MD trajectory of membrane buckle (POPC lipids) with embedded K160-Ub and K264-Ub RHDs. The movie covers frames from 15 to 25 μs showing trans-Ub-Ub contacts of the RHDs and dimer formation on the highly curved region of the buckle.

Supplementary Video 5

Coarse-grained MD trajectory of tubule deformation by five K160-Ub and five K264-Ub RHDs embedded along a 98 x 12 nm POPC tubule. The RHDs form clusters and occupy the highly curved caps of the tubule, severely deforming it.

Supplementary Video 6

Coarse-grained MD trajectory of spontaneous clustering and bud formation induced by nine (K160+K264)-Ub RHDs embedded in an asymmetric POPC bilayer ($\Delta N = 300$ and $\alpha = 1.00$).

Supplementary Video 7

Astigmatism-based 3D super-resolution reconstruction of an autophagosomes positive to FAM134B (magenta) and LC3B-II (green). Exchange DNA-PAINT imaging was performed in immunolabeled cells using custom docking strand-labelled secondary antibodies and short ATTO655-labelled oligonucleotides for transient target binding. The 3D movie was generated using ViSP open-source software v1.0.

Source data

Source Data Fig. 1–5 and Source Data Extended Data Fig. 1, 4, 7–10

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

González, A., Covarrubias-Pinto, A., Bhaskara, R.M. *et al.* Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum. *Nature* **618**, 394–401 (2023). https://doi.org/10.1038/s41586-023-06089-2

- Received: 29 March 2022
- Accepted: 14 April 2023
- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06089-2

Share this article

Anyone you share the following link with will be able to read this content:

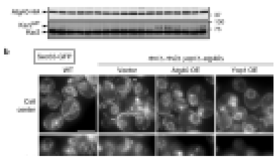
Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

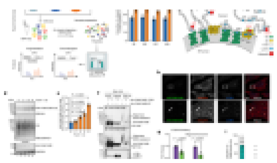
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06089-2



Super-assembly of ER-phagy receptor Atg40 induces local ER remodeling at contacts with forming autophagosomal membranes

03 July 2020

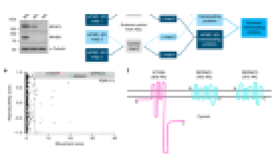
Keisuke Mochida, Akinori Yamasaki, … Hitoshi Nakatogawa



Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum

24 May 2023

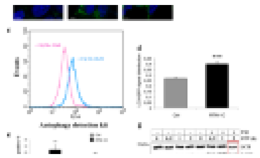
Alexis González, Adriana Covarrubias-Pinto, … Ivan Dikić



AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A

27 September 2018

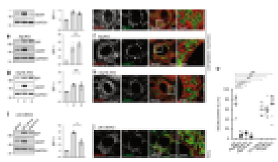
Alexandra K. Davies, Daniel N. Itzhak, … Georg H. H. Borner



Modulation of autophagy by RTN-1C: role in autophagosome biogenesis

18 November 2019

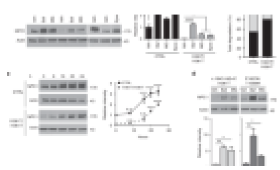
Manuela D’ Eletto, Anna Risuglia, … Federica Di Sano



ESCRT-III-driven piecemeal micro-ER-phagy remodels the ER during recovery from ER stress

07 November 2019

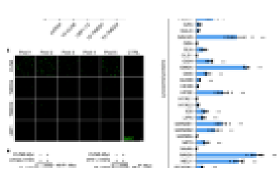
Marisa Loi, Andrea Raimondi, … Maurizio Molinari



Coordinate regulation of mutant NPC1 degradation by selective ER autophagy and MARCH6-dependent ERAD

10 September 2018

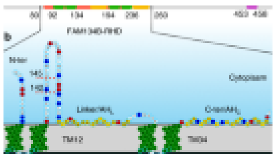
Mark L. Schultz, Kelsey L. Krus, … Andrew P. Lieberman



[CLN8 is an endoplasmic reticulum cargo receptor that regulates lysosome biogenesis]

05 November 2018

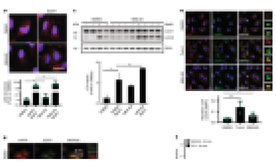
Alberto di Ronza, Lakshya Bajaj, … Marco Sardiello



[Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy]

30 May 2019

Ramachandra M. Bhaskara, Paolo Grumati, … Gerhard Hummer



[TRPML1 links lysosomal calcium to autophagosome biogenesis through the activation of the CaMKKβ/VPS34 pathway]

10 December 2019

A. Scotto Rosato, S. Montefusco, … D. L. Medina

- Article
- [Open Access]
- [Published: 24 May 2023]

Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy

- Hector Foronda,
- Yangxue Fu,
- Adriana Covarrubias-Pinto,
- Hartmut T. Bocker,
- Alexis González,
- Eric Seemann,
- Patricia Franzka,
- Andrea Bock,
- Ramachandra M. Bhaskara,
- Lutz Liebmann,
- Marina E. Hoffmann,
- Istvan Katona,
- Nicole Koch,
- Joachim Weis,
- Ingo Kurth,
- Joseph G. Gleeson,
- Fulvio Reggiori,
- Gerhard Hummer,
- Michael M. Kessels,
- Britta Qualmann,
- Muriel Mari,
- Ivan Dikić &
- …
- Christian A. Hübner

Show authors

Nature **volume 618**, pages 402–410 (2023)

- 6233 Accesses

- 1 Citations
- 193 Altmetric
- Metrics details

Abstract

Membrane-shaping proteins characterized by reticulon homology domains play an important part in the dynamic remodelling of the endoplasmic reticulum (ER). An example of such a protein is FAM134B, which can bind LC3 proteins and mediate the degradation of ER sheets through selective autophagy (ER-phagy)[1]. Mutations in *FAM134B* result in a neurodegenerative disorder in humans that mainly affects sensory and autonomic neurons[2]. Here we report that ARL6IP1, another ER-shaping protein that contains a reticulon homology domain and is associated with sensory loss[3], interacts with FAM134B and participates in the formation of heteromeric multi-protein clusters required for ER-phagy. Moreover, ubiquitination of ARL6IP1 promotes this process. Accordingly, disruption of *Arl6ip1* in mice causes an expansion of ER sheets in sensory neurons that degenerate over time. Primary cells obtained from *Arl6ip1*-deficient mice or from patients display incomplete budding of ER membranes and severe impairment of ER-phagy flux. Therefore, we propose that the clustering of ubiquitinated ER-shaping proteins facilitates the dynamic remodelling of the ER during ER-phagy and is important for neuronal maintenance.

Main

In a previous study[2], we identified *FAM134B* loss-of-function mutations in patients with autosomal recessive hereditary sensory and autonomic neuropathy (HSAN)[2]. This disorder is characterized by the degeneration of sensory and autonomic neurons that leads to numbness and the inability to feel pain. These symptoms in turn cause severe injuries and tissue damage[2]. Our studies further showed that FAM134B is an ER-resident membrane-shaping protein that can bind LC3 proteins and mediate the engulfment of parts of ER sheets by autophagosomes and their subsequent lysosomal

degradation[1]. The neurodegeneration that occurs in humans is also observed in *Fam134b*-deficient mice[1]. Together with the consequences observed following the disruption of the functional counterpart of FAM134B in yeast (Atg40)[4], these results suggest that the role of ER-phagy in cell viability is evolutionarily conserved.

Within the past decade, several other ER-resident membrane-shaping proteins with central reticulon homology domains (RHDs) have been associated with similar neurodegenerative disorders, including ATL1, ATL3, REEP1, REEP2, SPAST, RTN2, ARL6IP1 and LNPK[5,6,7]. Mutations in *ARL6IP1* cause SPG61, a neurodegenerative disorder characterized by progressive leg spasticity (hereditary spastic paraplegia (HSP)) in combination with loss of sensory and pain perception, thus overlapping with typical symptoms of HSAN[2,3,8]. The underlying mechanisms, however, remained largely elusive. Here we show that membrane-embedded clusters of ubiquitinated ARL6IP1 and FAM134B are required for effective ER remodelling and ER-phagy, defects of which result in severe neurodegeneration.

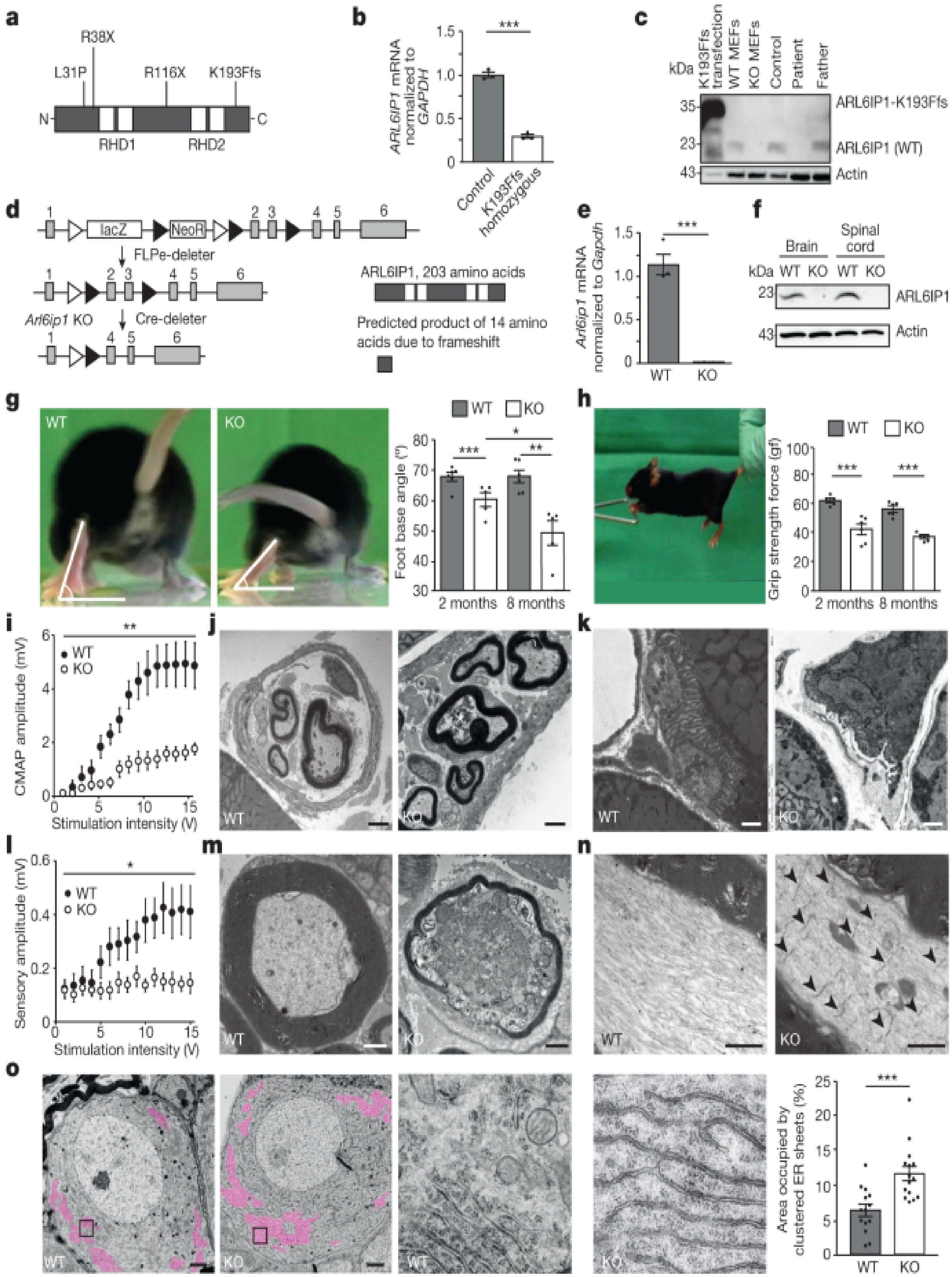
Degeneration of neurons in *Arl6ip1* knockout mice

To resolve the pathophysiology of the ARL6IP1-related disorder, we studied fibroblasts obtained from a patient with SPG61. This patient harboured the homozygous carboxy-terminal frameshift mutation *ARL6IP1* c.577–580delAAAC (NCBI Nucleotide database identifier NM_015161.3; K193Ffs variant) (Fig. 1a). Fibroblasts obtained from the patient's father, who was unaffected and a heterozygous carrier, and from an unrelated healthy individual (as a control) were also analysed. In silico analysis indicated that this frameshift mutation is predicted to result in the replacement of the 11 amino acids of the C terminus by 36 alternative residues[3]. Suggesting nonsense-mediated decay in vivo, *ARL6IP1* transcripts were absent in the patient's cells, as assessed by real-time PCR with exon spanning primers annealing 5′ end of the deletion (Fig. 1b). Immunoblot analyses of lysates from cells transfected with a plasmid encoding the K193Ffs variant produced a band with a slightly higher molecular weight (Fig. 1c). No variant protein was detected in fibroblasts obtained from the patient's father or from the patient when using an antibody directed against

the cytoplasmic loop of ARL6IP, a result that is in agreement with nonsense-mediated decay (Fig. 1c). Because the *ARL6IP1* c.577–580delAAAC variant represents a knockout (KO) allele, we generated *Arl6ip1* KO mice to model the ARL6IP1-associated disorder (Fig. 1d). KO of the gene was confirmed by both quantitative PCR (qPCR) of RNA isolated from mouse embryonic fibroblasts (MEFs) (Fig. 1e) and immunoblot analyses of MEFs and tissue lysates (Fig. 1c,f). Compared with wild-type (WT) mice, KO animals did not gain appropriate body weight (Extended Data Fig. 1a). Consistent with the CNS-related phenotypes reported for some patients[8,9], the weight of the brain in *Arl6ip1* KO mice was significantly decreased (0.46 g in WT mice compared with 0.39 g in KO mice at 6 months of age; $n = 3$, unpaired Student's t-test, $P = 0.016$). Moreover, cortical neuron (Extended Data Fig. 1b) and Purkinje cell (Extended Data Fig. 1c) counts were reduced. As a correlate of the HSP-related gait disorder, we measured the foot base angle of the hind paw at the moment when the toe was lifted[10] and found that it decreased with age in KO mice (Fig. 1g). Moreover, the grip strength of the upper limbs was reduced in KO mice (Fig. 1h). As reported for deceased patients with HSP, we observed that some axons connecting cortical and spinal cord motor neurons were swollen and full of dysfunctional organelles (Extended Data Fig. 1d). Because some patients develop muscle hypotonia and weakness, we also quantified spinal cord motor neurons, which progressively decreased in KO mice (Extended Data Fig. 1e). In agreement, the electrophysiological analysis showed a severe reduction in compound muscle action potential amplitudes (Fig. 1i). Consistent with neurogenic muscle atrophy, we found reduced musculus gastrocnemius mass (Extended Data Fig. 1f), degenerating grouped skeletal muscle fibres (Extended Data Fig. 1g) as well as degenerating intramuscular nerve fibres (Fig. 1j) and motor end plates (Fig. 1k and Extended Data Fig. 1h) in KO animals. Immunoblot analyses of brain protein lysates showed that the abundance of some ER-resident proteins with RHDs was altered in KO mice (Extended Data Fig. 1i). Sensory fibres were also degenerated, as evidenced by the substantial loss of sensory amplitudes in electrophysiological analyses of peripheral nerves (Fig. 1l). This result correlates with sensory loss and loss of pain perception in patients[3,8]. The ultrastructural analysis of peripheral nerves also showed swollen axons full of dysfunctional organelles and tubulofilamentous material (Fig. 1m), and ladder-like expansions of transverse ER sheet structures (Fig. 1n). This result is similar to that reported

previously for mice mutant for both ALT1 and REEP1 (ref. [11]). An analysis of cell bodies of peripheral sensory neurons (dorsal root ganglia (DRG)) uncovered a substantial expansion of ER sheets in *Arl6ip1* KO mice (Fig. 1o). Because of these observations and similar phenotypes in patients, we propose that ARL6IP1 plays a part in FAM134B-dependent ER-phagy.

Fig. 1: Neurodegeneration with ER sheet expansion in *Arl6ip1* KO mice.



a, Schematic of disease-associated ARL6IP1 variants. **b**, No *ARL6IP1* transcripts were detected in fibroblasts obtained from a patient carrying a

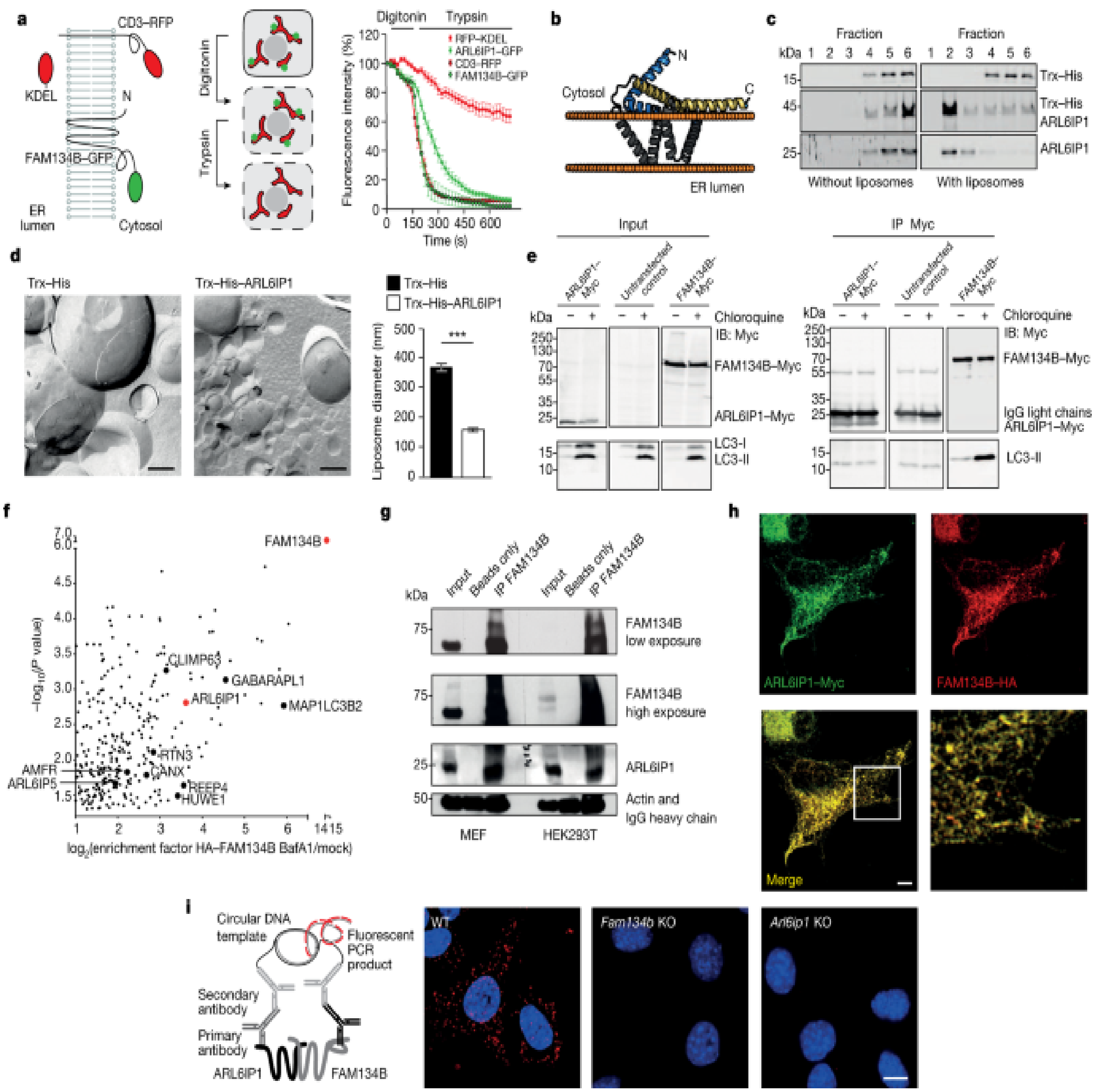
homozygous mutation of *ARL6IP1*^{K193Ffs} (qPCR with three replicates, two-sided unpaired Student's *t*-test, $P = 0.0001$). **c**, No variant protein was detected in cells from the patient (two experiments). **d**, *Arl6ip1* KO strategy. Frt sites, white triangles; *loxP* sites, black triangles. Predicted 14 amino acid product for the KO allele. **e**, No *Arl6ip1* transcripts were detected in KO MEFs (qPCR with three replicates, two-sided unpaired Student's t-test, $P = 0.0007$). **f**, Absence of ARL6IP1 in KO tissue lysates (two experiments). **g**, Decreasing foot base angle in KO mice ($n = 6$ WT mice and $n = 5$ KO mice; two-sided unpaired Student's *t*-test, 2 month-old WT versus KO, $P = 0.0007$; 8-month-old WT versus KO, $P = 0.0016$; KO 2 months versus 8 months, $P = 0.038$). **h**, Diminished forelimb grip strength in 2-month-old ($P = 0.0004$) and 8-month-old ($P = 0.0001$) KO mice ($n = 6$ WT and KO mice each; two-sided unpaired Student's *t*-test). **i**, Decreased compound muscle action potentials (CMAPs) in KO mice ($n = 6$ WT mice and $n = 7$ KO mice; repeated-measures analysis of variance (ANOVA), $F = 18.6$, $P = 0.0015$). **j**, TEM of intact WT and degenerating KO intramuscular nerve fibre (indicated by the asterisk). **k**, TEM of intact WT and degenerating KO motor end plate. **l**, Decreased sensory amplitudes in KO mice ($n = 6$ WT mice and $n = 7$ KO mice; repeated-measures ANOVA, $F = 6.08$, $P = 0.0314$). **m**, TEM of transversely cut intact WT and degenerating KO sciatic nerve axon. **n**, TEM of longitudinally cut sciatic nerve axons with ladder-like transverse ER sheet expansions (arrowheads) in KO but not WT mice. For **i**–**n**, analyses were performed using 6-month-old mice. **o**, TEM images (left) and quantification (right) show ER sheet expansions in lumbar spinal ganglion neurons in 12-month-old KO mice. ER sheet areas are coloured and higher magnifications are indicated. $n = 100$ cells from $n = 3$ WT mice and $n = 177$ cells from $n = 3$ KO mice; two-sided unpaired Student's *t*-test, $P = 0.0006$. Data shown as the mean ± s.e.m. Scale bars, 1 µm (**j**,**m**,**n**), 2 µm (**k**) or 2.5 µm (**o**).

Source Data

ARL6IP1 is part of ER-phagy complexes

On the basis of previous data[12] and results of fluorescence protease protection assays, we concluded that ARL6IP1 is characterized by the presence of RHD-like structural elements. That is, two long hydrophobic regions (transmembrane helical hairpins TM1+2 and TM3+4) separated by an accessible linker segment with both the amino and C termini facing the cytoplasm (Fig. 2a). The predicted structural model (produced using AlphaFold) of ARL6IP1 contains two membrane-embedded helical hairpins (TM1+2 and TM3+4) with two amphipathic helices (Fig. 2b and Extended Data Fig. 2a–c). Purified ARL6IP1 bound to liposomes in vitro (Fig. 2c) and increased the proportion of smaller liposomes (Fig. 2d), similar to FAM134B[1,13,14]. Despite predicted putative LC3-interacting regions (LIRs) in the N terminus or the cytoplasmic loop between RHDs, and in contrast to FAM134B, ARL6IP1 did not bind LC3 proteins (Fig. 2e), which suggested that ARL6IP1 is not an ER-phagy receptor on its own. However, ARL6IP1 was detected as an interaction partner in both a yeast two-hybrid screen for FAM134B-binding proteins[1] and in a proteomics analysis of FAM134B interactors (Fig. 2f). This result indicates that ARL6IP1 may be indirectly linked to the autophagy machinery through FAM134B. A sequence alignment analysis showed that both proteins are closely related and share all the signature membrane remodelling elements (Extended Data Fig. 2c). The co-precipitation of endogenous FAM134B and ARL6IP1 in MEFs or HEK293T cells (Fig. 2g) or endogenous FAM134B and expressed haemagglutinin (HA)-tagged ARL6IP1 (HA–ARL6IP1) in U2OS cells (Extended Data Fig. 2d) confirmed the interaction between both proteins. In agreement, ARL6IP1–Myc and FAM134B–HA co-localized to the same ER regions in MEFs (Fig. 2h). The co-localization results from MEFs was supported by data from proximity ligation assays using antibodies directed against endogenous FAM134B and ARL6IP1 (Fig. 2i). Using different deletion variants of ARL6IP1 and FAM134B, it became evident that their central parts containing the RHDs are required for the interaction between both proteins (Extended Data Fig. 2e–g). Notably, tagged FAM134A and FAM134C, the two homologues of FAM134B, also co-immunoprecipitated with ARL6IP1 in HEK293T cells (Extended Data Fig. 2e).

Fig. 2: FAM134B interacts with ARL6IP1.



a, Left, cytosolic location of the ARL6IP1 C terminus. Middle, COS-7 cells transiently expressing RFP-tagged or GFP-tagged ER proteins subjected to fluorescence protease protection assays by sequential administration of digitonin to permeabilize plasma but not intracellular membranes and then trypsin. Right, C-terminal fluorescent tags of ARL6IP1, FAM134B or CD3 are destroyed by trypsin, whereas luminal RFP–KDEL is protected (three experiments with $n = 29$ (CD3–RFP), 16 (RFP–KDEL), 15 (FAM134B–GFP) and 22 (ARL6IP1–GFP) cells). **b**, Three-dimensional model of ARL6IP1 built using AlphaFold2 showing the relative organization of key structural elements (grey and yellow) and their relative orientation in a model bilayer (orange beads). **c**, Recombinant Trx–His–ARL6IP1 and

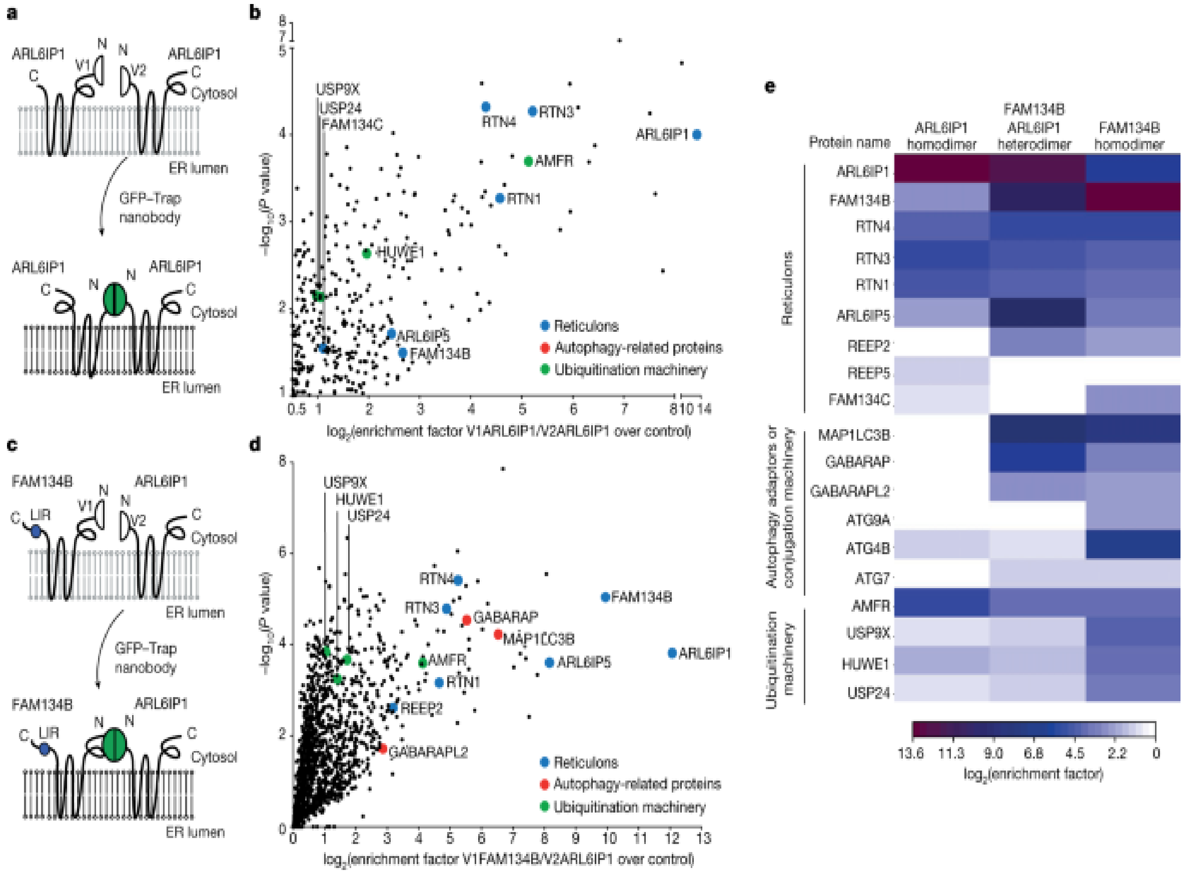
untagged ARL6IP1 float with lipid membranes in sucrose density gradients to the liposome fraction 2 (immunoblot analysis, $n = 2$). **d**, Left, TEM images of freeze-fractured incubations of liposomes with Trx–His or recombinant Trx–His–ARL6IP1. Right, mean liposome diameters are decreased with Trx–His–ARL6IP1 (1 experiment with $n = 1,817$ (Trx–His; average 380 nm) and 1,824 (Trx–His–ARL6IP1; average 150 nm) vesicles analysed; two-sided Mann–Whitney *U*-test, $P = 0.0001$). **e**, HEK293T cells were transfected with the indicated constructs and immunoblotted (IB). Pull-down with anti-Myc coupled beads shows that LC3-II co-precipitates with FAM134B–Myc but not ARL6IP1–Myc ($n = 1$). **f**, The FAM134B interactome in U2OS cells includes ARL6IP1 (single-sided volcano plot). Notable hits with a $\log_2$(enrichment factor) > 1 and $-\log_{10}$(*P* value) > 1.3 are highlighted (one-sided paired Student's *t*-test with three biological replicates). **g**, Co-precipitation of endogenous ARL6IP1 and FAM134B from MEFs and from HEK293T cells ($n = 1$). IP, immunoprecipitation. **h**, Overexpressed ARL6IP1–Myc and FAM134B–HA co-localize in MEFs. **i**, Proximity ligation assays suggest a proximity of less than 40 nm between endogenous ARL6IP1 and FAM134B. Specificity was confirmed by the absence of signals in the respective KO MEFs. Quantitative data are shown as the mean ± s.e.m. Scale bars, 200 nm (**d**) or 5 μm (**h**,**i**).

Source Data

We next used biomolecular complementation affinity purification (BiCAP)[15] to study the interactions of ARL6IP1 and FAM134B in vivo (Fig. 3a,c). For this purpose, ARL6IP1 and FAM134B were linked to the V1 or V2 segment of the fluorescent Venus protein. The Venus signals for V1–ARL6IP1 and V2–ARL6IP1 homodimers and for V1–FAM134B and V2–ARL6IP1 heterodimers were distributed along the ER. By contrast, no Venus signal was observed for the non-canonical ER-phagy receptor CCPG1 (ref. [16]) or for the ARL6IP1 variant lacking TM1 and TM2 (Extended Data Fig. 3a), which did not co-precipitate with FAM134B (Extended Data Fig. 2f). To characterize the functional relevance of this interaction, we immunoprecipitated homodimers and heterodimers and analysed interacting proteins by liquid chromatography and mass spectrometry (LC–MS)[17] (Fig. 3b,d,e). Within all interacting proteins, 7% exclusively interacted with ARL6IP1 homodimers, 52.4% exclusively with FAM134B homodimers and

approximately 40% with both (Extended Data Fig. 3b). Among the top ten gene ontology terms identified for interaction partners of ARL6IP1 homodimers were ER structural components (Extended Data Fig. 3c) such as FAM134B and FAM134C as well as the RHD proteins RTN1, RTN3 and RTN4 (Fig. 3b,e and Extended Data Fig. 3c,e). Notably, the main non-neuronal RTN4 isoform RTN4B was previously identified to interact with FAM134C and to have a role in autophagy[18]. Components of the autophagic vesicle formation machinery such as LC3B and GABARAPL2 were only identified as binding partners of ARL6IP1–FAM134B heterodimers but not ARL6IP1 homodimers (Fig. 3b,d,e and Extended Data Fig. 3c,d). Both ARL6IP1 and FAM134B homodimers interacted with components of the ubiquitination machinery (Fig. 3b,d,e). This included different E3 ligases, such as the ER-resident E3 ligase AMFR (also known as gp78) and HUWE1, and the deubiquitinating enzymes USP9X and USP24. Taken together, our data suggest that heterocomplexes of ARL6IP1 and FAM134B are part of ER-phagy receptor clusters.

Fig. 3: ARL6IP1 supports the formation of a multi-receptor complex with FAM134B and other ER-shaping proteins.



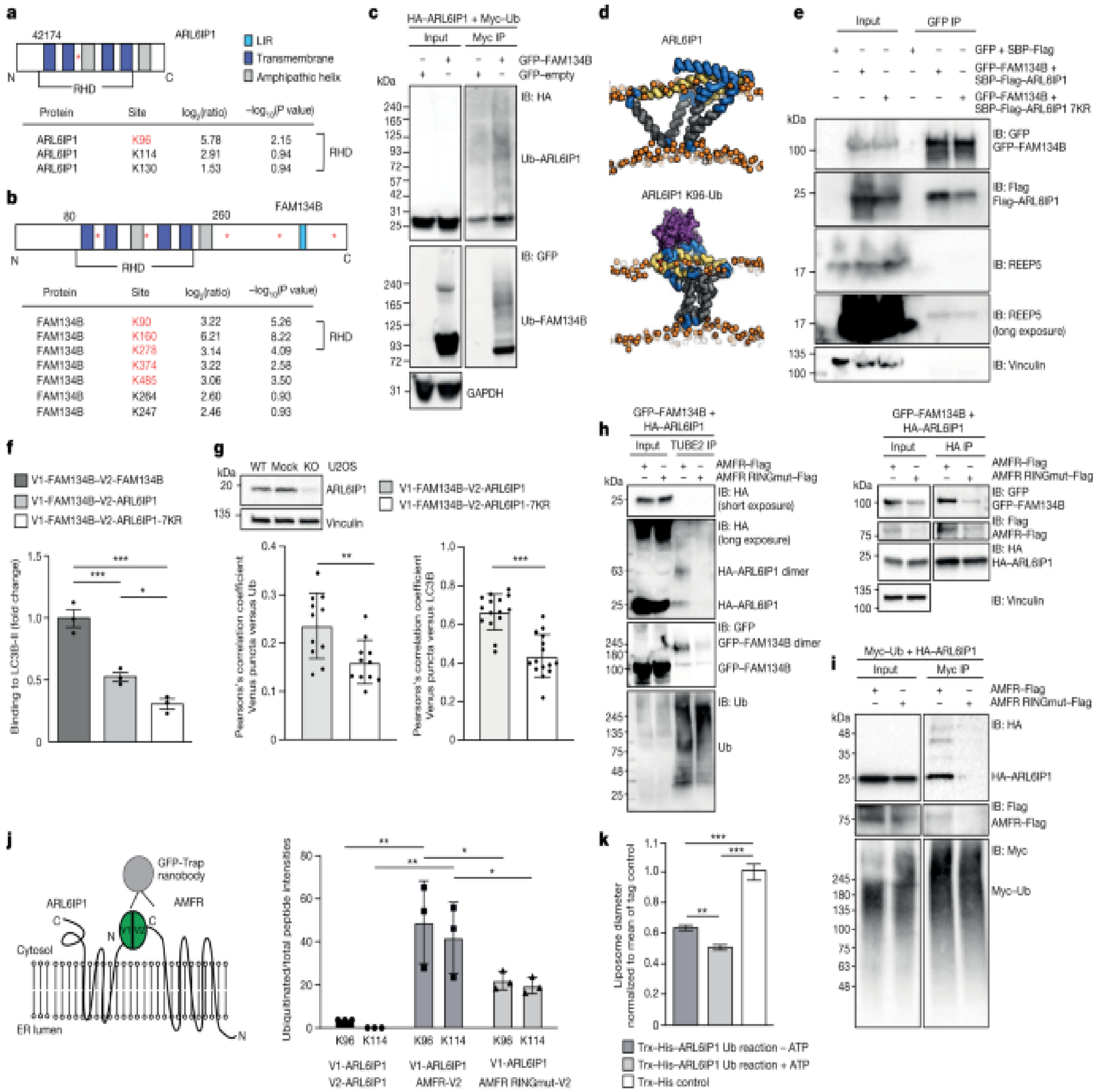
a, ARl6IP1 N-terminally tagged with the non-fluorescent N-terminal (V1) and ARl6IP1 N-terminally tagged with the C-terminal (V2) fragment of the Venus protein exhibit fluorescence after interaction. **b**, Single-sided Volcano plot of the label-free interactome of ARL6IP1 homodimers revealed that FAM134B, FAM134C and other RHD-containing ER proteins (blue) and proteins of the ubiquitination machinery (green) are binding partners. Only notable hits with $\log_2(P \text{ value}) > 1$ and $-\log_{10}(P \text{ value}) > 1.3$ are labelled in colour (one-sided paired Student's *t*-test with three biological replicates). **c**, FAM134B N-terminally tagged with the non-fluorescent N-terminal (V1) and ARL6IP1 N-terminally tagged with the C-terminal (V2) fragment of the Venus protein exhibit fluorescence after interaction. **d**, Single-sided volcano plot of the label-free interactome for V1-FAM134B–V2-ARL6IP1 heterodimers include autophagy-related proteins LC3B and GABARAPL2 (red) and FAM134B as known binding partners. Notable hits with $\log_2(P \text{ value}) > 1$ and $-\log_{10}(P \text{ value}) > 1.3$ are labelled in colour (one-sided paired Student's *t*-test with three biological replicates). Red dots indicate autophagy-related proteins. **e**, Heatmap of $\log_2$ enrichment of V1-ARL6IP1–

V2-ARL6IP1, V1-FAM134B–V2-ARL6IP1 or V1-FAM134B–V2-FAM134B over mock from notable hits indicated in **b** and **d**.

Ubiquitination promotes LC3B binding of FAM134B

As AMFR mediates the ubiquitination of FAM134B[17], we proposed that ubiquitination may also participate in the regulation of ARL6IP1. The LC–MS analysis identified several ubiquitinated lysine residues within ARL6IP1 in both ARL6IP1 homodimers and ARL6IP1–FAM134B heterodimers (Fig. 4a and Extended Data Fig. 4a,c–e). Most of the residues were located close to the RHDs of ARL6IP1, with K96 being significantly ubiquitinated and K114 and K130 potentially ubiquitinated (Fig. 4a). FAM134B exhibited an even higher number of ubiquitinated lysine residues, namely K90, K160, K278, K374 and K485 and potentially K247 and K264, which are also located close to its RHDs[17] (Fig. 4b and Extended Data Fig. 4b,f–l). Co-immunoprecipitation assays following the overexpression of HA–ARL6IP1 and Myc–ubiquitin in HEK293T cells confirmed that a significant amount of ARL6IP1 is ubiquitinated in the presence of FAM134B (Fig. 4c).

Fig. 4: LC3B binding to FAM134B–ARL6IP1 complexes depends on ARL6IP1 ubiquitination.



a,**b**, Position and $\log_2$(fold changes) of ubiquitinated lysine residues identified by LC–MS in ARL6IP1 (cells expressing V1-ARL6IP1–V2-ARL6IP1 or V1-FAM134B–V2-ARL6IP1) (**a**) and FAM134B (cells expressing V1-FAM134B–V2-FAM134B or V1-FAM134B–V2-ARL6IP1) (**b**). Significant sites are indicated in red ($-\log_{10}$(P value) > 1.3, one-sided unpaired Student's t-test). **c**, Anti-Myc IP from HEK293T cells co-expressing HA–ARL6IP1, GFP–FAM134B and Myc–ubiquitin (Ub) confirms ARL6IP1 and FAM134B ubiquitination (one experiment). **d**, Snapshots from coarse-grained molecular dynamics simulations showing the most populated conformations of non-ubiquitinated and ubiquitinated ARL6IP1 (ubiquitin purple) in 1-palmitoyl-2-oleoyl-glycero-3-

phosphocholine (POPC 16:0/18:1) bilayers (orange). **e**, Interaction with FAM134B but not endogenous REEP5 is promoted by ARL6IP1 ubiquitination (two experiments). **f**, LC3B co-precipitation with V1-FAM134B–V2-ARL6IP1-7KR is reduced compared with V1-FAM134B–V2-ARL6IP1 (three experiments; one-way ANOVA, $F = 46$; Holm–Sidak's post-test, FAM134B–FAM134B versus FAM134B–ARL6IP1, $P = 0.0014$; FAM134B–FAM134B versus FAM134B–ARL6IP1-7KR, $P = 0.0002$; FAM134B–ARL6IP1 versus FAM134B–ARL6IP1-7KR, $P = 0.0242$). **g**, Pearson's correlation analysis for ubiquitin-positive and LC3B-positive Venus puncta for V1-FAM134B–V2-ARL6IP1 and V1-FAM134B–V2-ARL6IP1-7KR in *ARL6IP1* KO U2OS cells (1 experiment with $n = 11$ (ubiquitin) and 15 (LC3B) cells; two-sided unpaired Student's *t*-test, $P = 0.006$ (ubiquitin) and $P = 0.0001$ (LC3B)). **h**, Left, ubiquitinated ARL6IP1 is detected after TUBE2 pull-down after co-expression of GFP–FAM134B and HA–ARL6IP1 with AMFR–Flag but not AMFR-RINGmut–Flag. Right, ARL6IP1–FAM134B interaction is promoted by AMFR but not AMFR-RINGmut–Flag ($n = 1$). **i**, Myc pull-down after co-expression of Myc–ubiquitin–HA–ARL6IP1 with AMFR–Flag or AMFR-RINGmut–Flag confirms ARL6IP1 ubiquitination with active AMFR ($n = 1$). **j**, Schematic (left) and quantification (right) of LC–MS of GFP pull-downs of co-expressed V1-ARL6IP1 and AMFR-V2 or AMFR-RINGmut-V2: AMFR ubiquitinates ARL6IP1 at K96 and K114 (one-sided unpaired Student's *t*-test: V1-ARL6IP1–V2-ARL6IP1 versus V1-ARL6IP1–AMFR-V2 K96, $P = 0.0075$; K114, $P = 0.006$; V1-ARL6IP1–AMFR-V2 versus V1-ARL6IP1–AMFR-RINGmut-V2 K96, $P = 0.038$; K114, $P = 0.044$). **k**, In vitro ubiquitination of Trx–His–ARL6IP1 with AMFR decreases the mean liposome diameter in TEM of freeze-fractured liposomes by about 20%. Incubations without ATP served as control (2 experiments with $n = 393$, 346 and 223 vesicles analysed (left to right); two-sided Kruskal–Wallis test with Dunn's post-hoc test: Trx–His–ARL6IP1 and ATP–Trx–His–ARL6IP1 + ATP, $P = 0.0024$; Trx–His–ARL6IP1 – ATP and control, $P < 0.0001$; Trx–His–ARL6IP1 + ATP and control, $P < 0.0001$). Quantitative data shown as the mean ± s.e.m.

Source Data

We next simulated the structural dynamics of non-ubiquitinated ARL6IP1 and the ubiquitinated ARL6IP1 (K96-Ub) embedded in phosphocholine

bilayers using coarse-grained molecular dynamics simulations (up to 10 μs). Ubiquitination resulted in a more compact conformation in which the ubiquitin moiety interacted with the cytosolic loops (Fig. 4d). To assess the functional relevance of ARL6IP1 ubiquitination, we replaced all seven predicted ubiquitinated lysine residues with arginine residues (ARL6IP1-7KR). Transiently expressed HA–ARL6IP1-7KR co-precipitated with green fluorescent protein (GFP)-tagged FAM134B (GFP–FAM134B), albeit to a lesser extent (Fig. 4e). A reduced interaction was also evident from co-immunoprecipitation of endogenous FAM134B with HA-ARL6IP1-7KR (Extended Data Fig. 2f). HA-ARL6IP1-7KR also co-localized with the ER sheet marker CLIMP63 (ref. [19]) (Extended Data Fig. 5a) and with FAM134B–Myc (Extended Data Fig. 5b). Moreover, the shaping properties of ARL6IP1-7KR tagged with a thioredoxin–histidine tag (Trx–His–ARL6IP1-7KR) did not differ from Trx–His–ARL6IP1 in liposome-binding assays (Extended Data Fig. 5c). ARL6IP1-7KR therefore preserves basic functions of ARL6IP1.

Next, we cloned ARL6IP1-7KR and FAM134B constructs with the complementing V1 and V2 segments of the Venus protein. Precipitation with anti-GFP-beads showed an efficient pull-down of V1-FAM134B–V2-ARL6IP1 heterodimers with LC3B. By contrast, binding to LC3B was significantly reduced for complexes with the ubiquitination-deficient ARL6IP1-7KR variant (Fig. 4f and Extended Data Fig. 5d). Co-expression of V1-ARL6IP1-7KR and V2-ARL6IP1 or V1-FAM134B and V2-ARL6IP1 in *ARL6IP1* KO U2OS cells resulted in a regular distribution of the Venus signal in the ER (Extended Data Fig. 5e). Co-localization analysis using Pearson's correlation showed that Venus puncta co-labelling with ubiquitin or LC3B in *ARL6IP1* KO U2OS cells were reduced for ARL6IP1-7KR (Fig. 4g and Extended Data Fig. 5e).

Because the ER-resident E3 ligase AMFR was found in the interactome of V1-FAM134B–V2-ARL6IP1 complexes, we tested whether AMFR ubiquitinates ARL6IP1. We detected ARL6IP1 and FAM134B in the tandem ubiquitin binding entity (TUBE2) pull-down from cell lysates upon co-expression of GFP–FAM134B and HA–ARL6IP1 together with AMFR–Flag, but not with the catalytically inactive AMFR-C356G-H361A variant (AMFR RINGmut-Flag) (Fig. 4h, left). Moreover, the interaction between ARL6IP1 and FAM134B was promoted by AMFR (Fig. 4h, right).

ARL6IP1 was detected in the Myc pull-down assay after co-expression of Myc–ubiquitin and ARL6IP1 with active AMFR, but not with the inactive AMFR RINGmut variant (Fig. 4i). When V1-ARL6IP1 was expressed with AMFR-V2, the ratio of peptides ubiquitinated at K96 or K114 in comparison to total peptides was approximately doubled compared with co-expression with AMFR RINGmut-V2 (Fig. 4j). We further detected ubiquitinated endogenous ARL6IP1 after induction of ER stress (Extended Data Fig. 5f), and we verified that ARL6IP1-K96 was ubiquitinated by AMFR in vitro (Extended Data Fig. 5g).

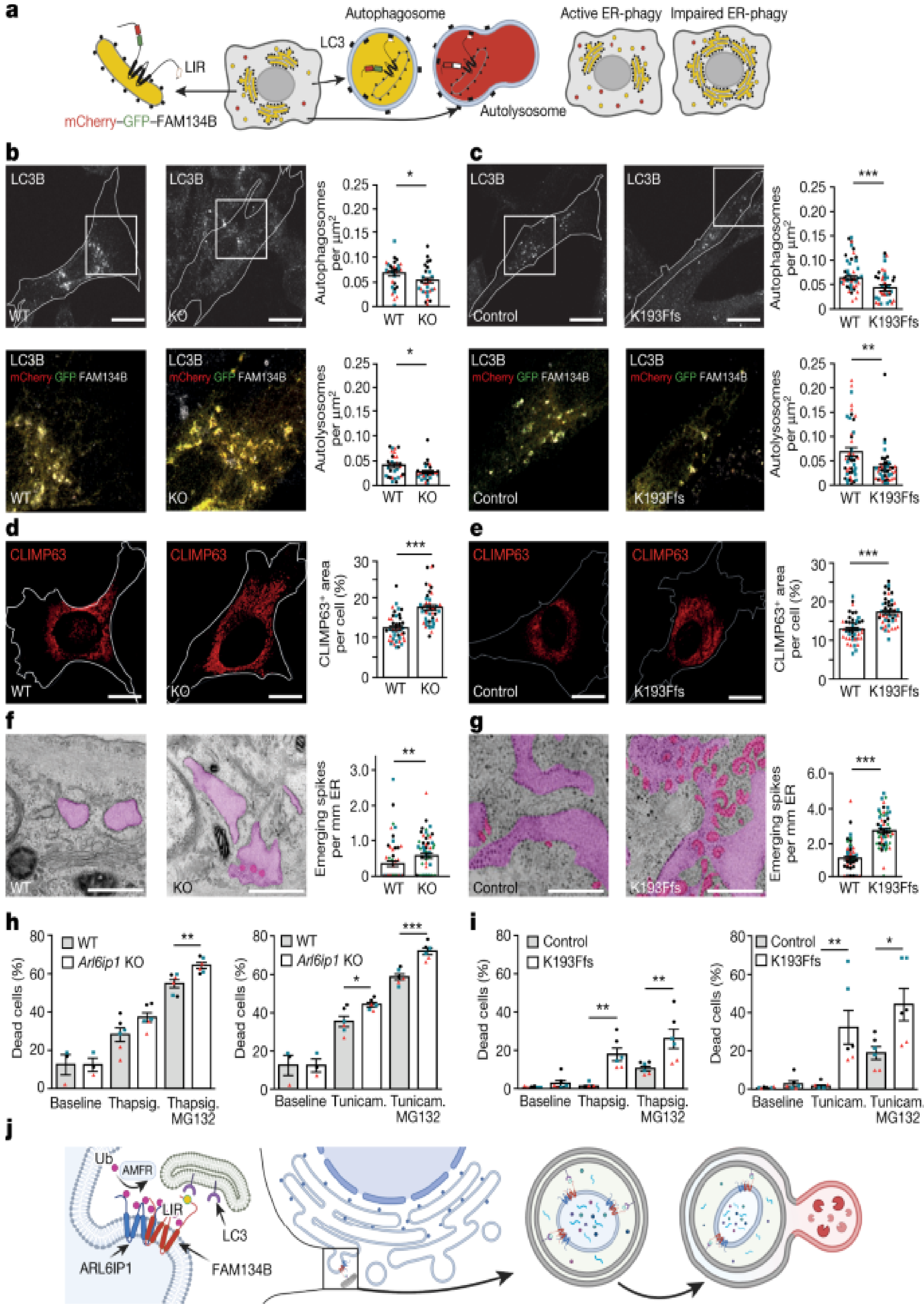
We considered that ubiquitination of ARL6IP1 might be important for membrane remodelling during ER-phagy. Accordingly, shaping assays showed that in vitro ubiquitination of ARL6IP1 by AMFR resulted in reduced mean liposome diameters (Fig. 4k). We therefore propose that the ubiquitination of RHDs of ARL6IP1 and FAM134B by AMFR is probably involved in ER remodelling during ER-phagy.

Impaired ER-phagy flux after ARL6IP1 disruption

To assess whether FAM134B-mediated ER-phagy is compromised in the absence of ARL6IP1, we overexpressed the mCherry–GFP–FAM134B reporter in WT and *Arl6ip1* KO MEFs (Fig. 5a). Staining for LC3B enabled the quantification of autophagosomes and autolysosomes because GFP is quenched in the acidic lumen of autolysosomes. Puncta labelled for LC3B, mCherry and GFP (autophagosomes) and puncta positive for LC3B and mCherry but negative for GFP (autolysosomes) were decreased in KO MEFs and in patient fibroblasts (Fig. 5b,c and Extended Data Fig. 6a). A defect in ER-phagy was further confirmed following siRNA-mediated knockdown of *ARL6IP1* in U2OS cells expressing the mCherry–GFP–FAM134B reporter after induction with doxycycline (Extended Data Fig. 6b,c). Similar results were also obtained after knockdown of *ARL6IP1* or *FAM134B* in HeLa cells alone or in combination, when the defect was more severe (Extended Data Fig. 7a,b). The quantification of SEC62-positive autophagosomes and autolysosomes in either starved *Arl6ip1* or *Fam134b* KO MEFs further confirmed a defect in ER-phagy. By contrast, SEC62-negative autophagosomes and autolysosomes as a readout for bulk autophagy were

not affected (Extended Data Fig. 7c). Intact bulk autophagy was additionally verified using the mCherry–GFP–LC3 reporter after knockdown of *ARL6IP1* in U2OS cells (Extended Data Fig. 7d,e). Mitophagy (Extended Data Fig. 7f) and pexophagy (Extended Data Fig. 7g) were not affected following the disruption of ARL6IP1. Notably, ER-phagy in *Arl6ip1* KO MEFs was rescued by ARL6IP1 or ARL6IP1-7KR overexpression (Extended Data Fig. 8).

Fig. 5: FAM134B-mediated ER-phagy requires ARL6IP1 in mice and humans.


a
mCherry–GFP–FAM134B
LIR
LC3
Autophagosome
Autolysosome
Active ER-phagy
Impaired ER-phagy
b
LC3B
WT
KO
Autophagosomes per μm²
Autolysosomes per μm²
mCherry GFP FAM134B
c
LC3B
Control
K193Ffs
WT
d
CLIMP63
CLIMP63⁺ area per cell (%)
e
f
Emerging spikes per mm ER
g
h
WT
Arl6ip1 KO
Dead cells (%)
Baseline
Thapsig.
Thapsig. MG132
Tunicam.
Tunicam. MG132
i
Control
K193Ffs
j
Ub
AMFR
LIR
LC3
ARL6IP1
FAM134B

a, Rational of the mCherry–GFP–FAM134B reporter. **b**,**c**, ARL6IP1 loss-of-function compromises ER-phagy. *Arl6ip1* WT and KO MEFs (**b**) or fibroblasts from the patient and the healthy individual (control) (**c**) were transfected with mCherry–GFP–FAM134B and stained for LC3B. Quantifications of $LC3B^{+}mCherry^{+}GFP^{+}$ puncta and $LC3B^{+}mCherry^{+}GFP^{-}$ puncta per cell area suggest that the formation of autophagosomes (3 experiments with 10 cells per genotype each; two-sided Mann–Whitney *U*-test; MEFs, $P = 0.0479$; human cells, $P = 0.0005$) and autolysosomes (3 experiments with 10 cells per genotype each; two-sided Mann–Whitney *U*-test; MEFs, $P = 0.0307$; human cells, $P = 0.0096$) is impaired. **d**,**e**, ARL6IP1 loss-of-function enlarges ER sheets. *Arl6ip1* WT and KO MEFs (**d**) or fibroblasts from the patient and healthy individual (**e**) were stained for the ER sheet protein CLIMP63 and the relative $CLIMP63^{+}$ area per cell calculated (3 experiments with 15 cells per genotype; two-sided Mann–Whitney *U*-test; MEFs, $P = 0.0001$; human cells, $P = 0.0001$). **f**,**g**, TEM images showed increased numbers of small highly curved ER protrusions emanating from ER sheets in the absence of ARL6IP1 (ER sheets in light purple, ER-emerging spikes in pink) in MEFs (**f**) and human fibroblasts (**g**) (1 experiment with 55 cells per genotype; two-sided Mann–Whitney *U*-test; $P = 0.0059$ (**f**) and $P = 0.0001$ (**g**)). **h**,**i**, ARL6IP1 loss-of-function decreases cell viability in the presence of the ER stressors tunicamycin (Tunicam.) or thapsigargin (Thapsig.) with or without the proteasome inhibitor MG132 in MEFs (**h**) and human fibroblasts (**i**) (3 experiments with 2 replicates; two-sided Mann–Whitney *U*-test; MEFs, thapsigargin + MG132, $P = 0.0022$; tunicamycin, $P = 0.041$; tunicamycin + MG132, $P = 0.0022$; patient cells, thapsigargin, $P = 0.0022$; thapsigargin + MG132, $P = 0.0087$; tunicamycin, $P = 0.0022$; tunicamycin + MG132, $P = 0.026$). **j**, Cartoon showing ARL6IP1 (blue) ubiquitinated (red balls) by AMFR in a complex with FAM134B (red), which binds to LC3 (purple) during ER-phagy. Quantitative data are shown as the mean ± s.e.m. Individual experiments are indicated by differently coloured data points. Scale bars, 500 nm (**f**,**g**), 10 µm (**d**,**e**) or 20 µm (**b**,**c**). The models in **a** and **j** were created using BioRender (https://www.biorender.com).

Source Data

Because FAM134B is involved in the lysosomal clearance of ER subdomains containing ERAD-resistant misfolded proteins such as pro-

collagen[20], we also assessed whether the clearance of pro-collagen is affected following disruption of ARL6IP1. The ratio of cells with an intracellular accumulation of pro-collagen was substantially increased in both *Fam134b* and *Arl6ip1* KO MEFs (Extended Data Fig. 9a). A defect in collagen exocytosis could be excluded because collagen concentrations in cell culture supernatants did not differ between genotypes (Extended Data Fig. 9b). Comparable results were obtained for fibroblasts from the patient and from the healthy individual (Extended Data Fig. 9c,d), which were confirmed by immunoblot analysis (Extended Data Fig. 9e,f). Co-staining with LAMP2 further indicated that the delivery of pro-collagen to lysosomes is impaired in *Fam134b* and *Arl6ip1* KO MEFs (Extended Data Fig. 9g).

Collectively, these data suggest that FAM134B-mediated ER-phagy requires ARL6IP1 and that it is likely to be promoted by ARL6IP1 ubiquitination.

Abnormal ER sheets after ARL6IP1 disruption

In agreement with a severe defect of FAM134B-dependent ER-phagy flux, the relative area covered by CLIMP63-labelled ER sheets was increased in maximum intensity projections of confocal *z*-stacks of *Arl6ip1* KO MEFs (Fig. 5d) and in patient fibroblasts (Fig. 5e). The area labelled for RTN4, a marker of ER tubules[21], was decreased, which closely resembled findings for *Fam134b* KO MEFs (Extended Data Fig. 10a).

Transmission electron microscopy (TEM) analyses revealed an increased number of spike-like ER protrusions in *Arl6ip1* KO MEFs (Fig. 5f). The numbers of spike-like ER protrusions was higher in fibroblasts from the patient compared with those from the healthy individual (Fig. 5g). Notably, such ER protrusions were also increased in fibroblasts from the patient's father (heterozygous for K193Ffs) compared with the healthy individual, but significantly less compared with the patient (Extended Data Fig. 10b).

The compromised ER structure was accompanied by diminished cell viability in response to ER stressors such as thapsigargin or tunicamycin with or without the proteasome inhibitor MG132 in KO MEFs (Fig. 5h) and in fibroblasts from the patient (Fig. 5i). By contrast, the viability of fibroblasts from the patient's father only differed marginally from control

cells (Extended Data Fig. 10c). Notably, the alterations in ER structure were more pronounced in human fibroblasts compared with MEFs, whereas the sensitivity was less for human fibroblasts in our cell viability assays. This disparity may be explained by species-specific differences or differences between embryonic and adult tissues. Overall, we showed that the disruption of ARL6IP1 leads to a defect in ER remodelling and ER structure and decreases resistance to ER stress.

We conclude that the loss of ARL6IP1 crucially impairs ER-phagy and decreases cellular fitness. As an in vivo correlate of defective ER-phagy, ER sheets were expanded in DRG neurons in *Arl6ip1* KO mice, which degenerated over time and led to sensory loss. In agreement, *ARL6IP1* knockdown was reported to result in the expansion of ER sheets in HeLa cells[22] and the fragmentation of ER tubules in *Drosophila* axons[23]. Here we showed that ARL6IP1 is required for FAM134B-mediated ER-phagy. However, it does not act as an ER-phagy receptor on its own (Fig. 5j) because it cannot bind LC3 proteins itself. Instead, ARL6IP1 forms multi-protein clusters with FAM134B and other ER-shaping proteins. Notably, the RHDs of FAM134B and ARL6IP1 were ubiquitinated within such ER-phagy competent multi-protein clusters. Moreover, the ubiquitination of ARL6IP1 increased its shaping properties and its binding to LC3 by FAM134B. Our observation that membrane protrusions emanating from ER sheets are increased suggests that ER membrane remodelling is incomplete in the absence of ARL6IP1. As a new mechanism, we propose that ER-phagy is controlled by the ubiquitin-dependent formation of heteromeric ER-phagy receptor complexes. Ubiquitination therefore acts as a higher level of regulation acting on top of RHD phosphorylation to promote FAM134B oligomerization and ER fragmentation[24]. As ER-phagy is also involved in the degradation of ER subdomains containing ERAD-resistant misfolded proteins, which has been shown for ATZ polymers[25] and for endogenous pro-collagen[20], compromised ER-phagy following the disruption of either FAM134B or ARL6IP1 coupled with age-associated attenuation of autophagy might contribute to an accumulation of misfolded or aggregated proteins within the ER. Subsequently, this leads to impaired proteostasis and progressive neurodegeneration. Because ARL6IP1 also binds to FAM134A and FAM134C, two broadly expressed close homologues of FAM134B[2], which also serve as ER-phagy receptors[26], this

may explain why mutations in *ARL6IP1* lead to more severe disease[2,3,8,9,27]. In light of the evolutionarily conserved function of FAM134B, the formation of heteromeric clusters of ubiquitinated membrane-shaping proteins to remodel the ER may represent a more general principle of cell homeostasis.

Methods

To study the function of ARL6IP1 and the consequences of its disruption, we investigated *Arl6ip1* KO mice and different cell culture models, including fibroblasts from a patient carrying the homozygous mutation $AR6IP1^{K193Ffs}$, from the patient's father (unaffected and carrying the heterozygous allele) and from a healthy individual. Mouse experiments were performed on a C57BL/6 background after backcrossing for more than four generations. Mice were maintained in groups of up to 3 mice per cage at 21 ± 2 °C, air humidity of ≥45%, 15-fold air exchange, 14–10 h day–night cycle and maximum 500 lx. Mice had free access to standard mouse chow and water. Littermates of the same sex were randomly assigned to experimental groups. Experiments were conducted in a blinded manner with regard to cell, mouse and human genotypes. Figure legends include details of replicate experiments used to generate datasets. All animal experiments were approved by the Thüringer Landesamt für Lebensmittelsicherheit und Verbraucherschutz (registration numbers 02-055/14 and UKJ-17-006). Studies using human fibroblasts were approved by the local ethics committee.

Plasmids are presented in Supplementary Table 1. cDNAs were cloned into the pDONR223 vector using a BP Clonase Reaction kit (Invitrogen, 11789100) and further recombined into the Gateway destination vectors pcDNA5-FRT/TO-N-mCherry-EGFP, pcDNA3.1-N-HA, pHAGE-GFP, pcDNA3.1-N-Flag, pcDNA3.1-C-Flag, pcDNA3.1-N-SBP-Flag, pGEX6-GST, and the biomolecular complementation affinity purification system vectors pDEST-V1-ORF, pDEST-V2-ORF, pDEST-ORF-V1 and pDEST-ORF-V2 using a LR Clonase Reaction kit (Invitrogen). Plasmids encoding untagged ARL6IP1, Trx–His-tagged ARL6IP1 and Trx–His–ARL6IP1-7KR were generated by subcloning ARL6IP1 from pGEX6P1 into pPal7 and pET32a, respectively. ARL6IP1 in pGEX6P1 was cloned by PCR using

pClneo-ARL6IP1 WT-GFP as template. The three sgRNA guides of ARL6IP1 were cloned into the pLentiCRISPR v2 vector.

Primers are presented in Supplementary Table 2, primary antibodies in Supplementary Table 3 and secondary antibodies are presented in Supplementary Table 4.

Generation of *Arl6ip1* KO mice

The EUCOMM embryonic stem cell clone HEPD0752_7_A11 (Source Bioscience) was injected into C57BL/6J donor blastocysts. Next, 15–30-week-old F_1 female offspring from C57BL/6J and CBA/J matings served as foster mice. Resulting chimeras were mated with C57BL/6J. For all experiments, littermates were used, which had been backcrossed for at least four generations. Genotyping was performed by PCR with three primers (*Arl6ip1*-forward: 5′-GTAATATTCTGAGCACTGCCT-3′, *Arl6ip1*-KO-reverse: 5′-TGCCATAATGACCTAATACTGTTGTG-3′, *Arl6ip1*-WT-reverse: 5′-CTAAGCACAGGCTATGAACC-3), which produced a WT band of 537 bp or the KO band of 350 bp.

Generation of *ARL6IP1* CRISPR–Cas9 KO cell lines

The *ARL6IP1* knockout U2OS cell line was generated using a lentiviral CRISPR–Cas9 system. sgRNAs are reported in Supplementary Table 5 (design at https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). The lentiviral plasmids were generated as previously reported[28]. The forward and reverse oligonucleotides were annealed and phosphorylated using T4 polynucleotide kinase (BioLabs). The oligonucleotides were ligated into the Cas9 vector pLenti-Puro-v2 and pLenti-Puro-EGFP using the BsmBI site. The lentiviral plasmids were co-transfected into HEK293T cells together with the packaging vectors pPAX2 and pDM2.G for lentivirus production. After 48 h, the medium containing lentiviral particles was collected, centrifuged to remove dead cells and stored at –80 °C. To generate the *ARL6IP1* KO cell line, fresh U2OS or HeLa cells were infected with lentiviral particles with the three different sgRNAs for 48 h and then selected using 5 μg ml^{-1} of puromycin. The surviving cells were maintained in DMEM supplemented with 2 μg ml^{-1} puromycin. When using pLenti-Puro-

EGFP as backbone of the sgRNA, cells were also FACS-sorted for GFP expression (SONY SH800S Cell Sorter, version 2.1.6). KO was verified by western blotting.

Motor performance

For the beam walk test, mice were placed on an elevated beam of 1 m in length and 4 cm in width, with the home cage at the end. After habituation on three consecutive days, the mouse was videotaped from behind during its movement on the beam. The foot base angle of the hind limb was measured at the moment when the toe was lifted.

For the grip strength analysis, mice were lifted at the tail base, brought to a trapeze-shaped handle connected with a force sensor (Grip Strength Meter, Ugo Basile). When the mouse spontaneously grabbed the handle, the mouse was gradually pulled away from the handle until it was released.

For the electrophysiological analysis of peripheral nerves, anaesthetized mice (100 mg kg^{-1} ketamine and 16 mg kg^{-1} xylazine) were placed on a heating pad. One pair of needle electrodes with a tip distance of 5 mm (WE30030.1H10, Science Products) was inserted near the base of the tail and a second pair 30 mm distal to the stimulation site close to the tip of the tail. For the analysis of motor fibres, the stimulus was applied through the proximal electrodes and the response recorded using the distal electrodes. Compound muscle action potentials and sensory nerve action potentials were evoked with increasing intensity (0–15 V, increment of 1 V, 50 μs duration, interstimulus interval of 20 s). Sum action potentials were filtered (high-pass filter 3 Hz, low-pass filter 1.3 kHz) and digitized with a sampling frequency of 10 kHz. Amplitudes were determined from peak to peak.

Histology, neuron count and TEM of mouse tissues

Mice were deeply anaesthetized and transcardially perfused with PBS (pH 7.4) followed by 4% paraformaldehyde (PFA) in PBS for 10 min. After dissection, tissues were post-fixed in 4% PFA in PBS for at least 1 h. Tissues were incubated in sucrose (10% sucrose for 4 h and in 30% sucrose overnight at 4 °C), frozen on dry ice and cut with a sliding microtome (Leica

SM 2000R) in 30-µm-thick free-floating sections and stored in PBS supplemented with sodium azide at 4 °C until further use. For NeuN staining, free-floating sections were permeabilized with 0.25% Triton X-100, blocked in 5% normal goat serum (NGS) for 1 h and incubated with mouse anti-NeuN (Millipore) 1:500 at 4 °C overnight. After washing, sections were incubated with the appropriate secondary antibodies (Invitrogen) at 1:1,000 and Hoechst 33342 (Thermo Scientific). Sections were mounted with Fluoromount-G (Southern Biotech). Images were acquired using Cellobserver Z1 (Zeiss) with the tile-scan module and further analysed using ImageJ.

For paraffin embedding, the samples were dehydrated overnight in a series of ethanol and xylol baths (Leica TP20 Tissue Processor), embedded with paraffin (Leica HistoCore Arcadia) and cut into 5 µm sections with a microtome (ThermoScientific Microm HM 355S). For histological analyses, sections were either stained with haematoxylin and eosin (Sigma-Aldrich) or cresyl-violet (Sigma-Aldrich). Images were acquired using a Zeiss AxioLab A1 microscope and further analysed using ImageJ.

For TEM of tissue sections, animals were perfused with 2.5% glutaraldehyde in PBS unless indicated otherwise. For the analysis of DRGs, mice were perfused with 4% PFA and 2% glutaraldehyde in PBS. After dissection, tissues were post-fixed overnight. Tissues were contrasted with 1% osmium tetroxide, dehydrated and infiltrated with epoxy resin. Ultrathin sections were stained with uranyl acetate and lead citrate, mounted on copper grids and viewed with a Philips CM10 or Zeiss EM 900 digital (DRGs) transmission electron microscope.

TEM of cultured cells

MEFs and human fibroblasts were fixed by adding an equal volume of double strength fixative (4% PFA, 5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4) to the culture medium and incubated for 20 min at room temperature. The fixative mixture was then replaced with one volume of single strength fixative (2% PFA and 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4) for another 2 h at room temperature. After 5 washes with 0.1 M sodium cacodylate buffer (pH 7.4), cells were

processed for dehydration and embedding in Epon resin[29]. To preserve their original morphology, the monolayer culture of fibroblasts of the patient and of his father were embedded in their original position in their culture flasks. By contrast, WT and *Arl6ip1* KO MEFs were scraped into 2% low-melting-point agarose before the dehydration process and the embedding in Epon resin, as previously described[29]. Subsequently, 70 nm ultrathin sections were cut using a Leica EM UC7 ultra microtome (Leica Microsystems) and stained with uranyl acetate and lead citrate[29]. Cell sections were analysed using an 80 kV transmission electron microscope CM100bio TEM (FEI).

Skeletal muscle fibre bundle staining

Muscles freshly dissected from 2-month-old mice were fixed in 2% PFA for 15 min and subsequently washed with PBS. Fibre bundles were prepared and used for further analyses. After overnight permeabilization with 0.2% Triton X-100 in PBS, samples were blocked with 5% NGS for 1 h followed by an incubation with α-bungarotoxin-Alexa 555 (Invitrogen) 1:500 and mouse anti-NF200 overnight at 4 °C. After washing with PBS, single myofibre bundles were incubated with the corresponding secondary antibodies (Invitrogen) in a dilution of 1:1,000 for 1 h at room temperature. Nuclei were stained with Hoechst 33258 (Invitrogen). Myofibres were washed with PBS and mounted using Fluoromount-G (Southern Biotech). Images were acquired using a Zeiss LSM 880 confocal microscope with Airyscan using the *z*-stack module. *Z*-projections with maximum intensities processed using ImageJ are shown.

Cell culture

HEK293T, U2OS and HeLa cells were obtained from the American Type Culture Collection. Their identities were authenticated by STR analysis. U2OS TRex cells were provided by S. Blacklow (Brigham and Women's Hospital and Harvard Medical School), HeLa TRex were provided by S. Taylor (Manchester University). WT, *Fam134b* KO and *Arl6ip1* KO MEFs were isolated from embryos and immortalized using SV40 large T antigen. All cell lines were regularly tested for mycoplasma contamination using a LookOut Mycoplasma PCR Detection kit (Sigma). Cells were maintained at

37 °C with 5% CO_2 in DMEM medium (Gibco) supplemented with 10% FBS (Gibco) and 100 U ml^{-1} penicillin and streptomycin (Gibco).

Inducible cell lines were induced with 1 µg ml^{-1} doxycycline (Sigma-Aldrich).

Bafilomycin A1 (LC-Laboratories) was used at a concentration of 200 ng ml^{-1}, Torin1 (LC-Laboratories) at 250 nM. EBSS medium was obtained from Gibco. For each treatment, cells were plated the day before to perform the experiments when cells had a confluence of 50–60%. For transient expression, DNA plasmids were transfected using GeneJuice (Merck-Millipore), Turbofect (Thermo Scientific) or Lipofectamine 2000 (Invitrogen).

U2OS TRex cell lines were used to generate stable cell lines using the Flp-IN TRex system (Invitrogen) or the lentiviral vector p-Lenti N-HA. For *ARL6IP1* knockdown experiments, the respective cells were transfected with either 30 pmol siNT (non-targeting sequence, Qiagen) or with 30 pmol double-stranded *ARL6IP1* siRNA (Integrated DNA Technologies; hs.RiARL6IP1.13.2) using Lipofectamine RNAiMAX transfection reagent (Invitrogen, 13778075). For *ARL6IP1* and *FAM134B* double-knockdown experiments, HeLa cells were transfected with *ARL6IP1* siRNA and *FAM134B* siRNA (siRNA RETREG1 18 J-016936-18-0002 and siRNA RETREG1 21 J-016936-21-0002, respectively). Experiments were performed 72 h after transfection.

Protein isolation from cells and tissue lysates

Cells were collected and lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% (v/v) NP-40, 1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 1 mM EDTA and complete protease inhibitor (Roche)). Tissue lysates were prepared using a Ultra-Turrax T8 tissue homogenizer (IKA-WERKE) in RIPA buffer. After sonication, homogenates were spun down at 16,900*g* to remove nuclei and insoluble debris. The supernatant was stored at –80 °C until further use.

Western blotting

Proteins were denatured at 90 °C for 5 min in Laemmli buffer, resolved by SDS–PAGE and transferred to methanol-activated PVDF membranes (Amersham Hybond P 0.45 µm). Membranes were blocked for 1 h in 10% skim milk in TBS-T (Tris-buffered saline with Tween, 20 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.5) and incubated overnight at 4 °C with the specific primary antibody followed by 1 h incubation with the respective secondary antibody at room temperature. Detection was carried out using Clarity Western ECL substrate (Bio-Rad) and a LAS 4000 automated detection system (GE Healthcare). Bands were quantified using ImageJ.

Real-time qPCR

RNA was isolated by TRIzol–chloroform extraction. RNA was reverse-transcribed using a GoScript reverse transcription kit (Promega). qPCR was performed with a final amount of 20 ng of cDNA and EvaGreen Mix (Bio-Rad) with primer pairs for either mouse *Gapdh* (forward, GCTCATGACCACAGTCCAT; reverse, GTCATCATACTTGGCAGGTTT), mouse *Arl6ip1* (forward, GCTCTAATAAATGGACCACTG; reverse, GCACAAATGTCACAATCAGGT), human *GAPDH* (forward, GAAGGCTGGGGCTCATTT; reverse, GGACTGTGGTCATGAGTC) or human *ARL6IP1* (forward, GCTCCAATAAATGGACCACTGA; reverse, GGAAGTCACTATCAGGTAGGT) on a CFX96 Touch Real-Time PCR detection system (Bio-Rad).

Immunofluorescence and autophagic flux analysis

For immunostainings, cells were fixed with 4% PFA at room temperature or ice-cold methanol, washed with PBS and permeabilized with 0.25% Triton X-100 in PBS at room temperature for 10 min, or with 0.1% saponin in PBS at room temperature for 1 min followed by blocking with 5% NGS in PBS for 1 h at room temperature. Incubation with primary antibodies diluted in 5% NGS and 0.25% Triton X-100 in PBS was carried out overnight at 4 °C or at room temperature for 1 h. After three consecutive washes with PBS, secondary antibodies were incubated for 1 h at room temperature. After three washes, cells were incubated at room temperature with Hoechst 33258 (Invitrogen) for 10 min. After a final wash with PBS, the coverslips were

mounted with Fluoromount-G solution (ThermoFisher, 00-4958-02). Images were acquired using a confocal microscope (Zeiss LSM 880 with Airyscan).

For the analysis of ER-phagy, cells were transiently transfected with the mCherry–GFP–FAM134B reporter construct. Cells were fixed with ice-cold methanol for 10 min 24 h after transfection. Then cells were permeabilized with 0.25% Triton X-100 in PBS and blocked with 4% NGS for 1 h and stained for LC3B. Images were taken with a LSM 880 and analysed using the ComDet (v.0.5.5) plugin for ImageJ (https://github.com/ekatrukha/ComDet; settings: particle size = 10 pixels, co-localization distance = 7, intensity threshold = 20). The cell border was selected and the cell area determined. Only signals within the cell border were analysed. The intensity threshold was set at 1,000 for all channels, except for human fibroblasts, for which the threshold for mCherry was set at 200. Pearson's coefficients were calculated using the JaCOP plugin in ImageJ and normalized to the ER area.

To assess whether ARL6IP1 is involved in bulk autophagy, we knocked down *ARL6IP1* in U2OS cells stably expressing the mCherry–GFP–LC3 reporter. Autophagy was triggered by 6 h of EBSS exposure or 6 h of 250 nM Torin1 exposure. Images were acquired with a high-content microscope–Yokogawa CQ1 confocal imaging cytometer. To assess whether ARL6IP1 is involved in mitophagy, ARL6IP1, CRISPR–Cas9 KO HeLa cells were transfected with the mitophagy reporter mCherry–GFP–FIS1. Autophagic flux was triggered with 40 µM CCCP for 4 h. Pexophagy was assessed after siRNA-mediated knockdown of *ARL6IP1* in U2OS cells after induction of the doxycycline-inducible reporter mCherry–GFP–PMP34 (ref. [30]) at baseline or after starvation with EBSS for 20 h. Images were acquired with a confocal microscope (Zeiss LSM 880 with Airyscan). The red and yellow puncta were manually counted.

Rescue experiment

Arl6ip1 WT and KO MEFs were seeded in 24-well-plates at 40,000 cells per well. After 24 h, cells were transfected with the mCherry–GFP–FAM134B plasmid in combination with either the ARL6IP–HA or the ARL6IP1-7KR–HA plasmid. After 48 h, cells were fixed with ice-cold methanol for 10 min,

permeabilized with 0.25% Triton X-100 in PBS and blocked with 4% NGS for 1 h and stained for LC3B and HA and further processed as described above. Images were taken with a LSM 880 and analysed using the ComDet (v.0.5.5) plugin for ImageJ (https://github.com/ekatrukha/ComDet; settings: particle size = 10 pixels, co-localization distance = 10, intensity threshold = 200/200/15). Only signals within the cell border were analysed.

ER stress induction and cell viability count

MEFs or human fibroblasts were seeded in 6-well-plates and cultured to 70–80% confluency. Cells were washed with PBS and incubated with new medium with 1.5 μM thapsigargin (Sigma, T9033-5MG) or 5 μg ml^{-1} tunicamycin (Santa Cruz) without or in combination with 1 μM MG132 (Calbiochem, 474787-10MG) to inhibit the proteasome. After 24 h for MEFs and 48 h for human fibroblasts, the culture medium was removed, cells washed with PBS and trypsinized. All cells were pooled, centrifuged at 800 r.p.m. for 5 min (Heraeus Sepatech Megafuge 2.0R) and resuspended in fresh medium. Cell viability was measured by trypan blue exclusion with an automatic counting device (Bio-Rad TC20 automatic cell counter).

Interactome analysis and sample preparation for MS

For the LC–MS interactome analysis, HA–FAM134B expression was induced in U2OS cells with doxycycline. HEK293T cells were transiently transfected with the constructs V1-ARL6IP1, V2-ARL6IP1, V1-FAM134B and V2-FAM134B. For in vivo ubiquitination, HEK293T cells were transiently transfected with the plasmids V1-ARL6IP1, V2-ARL6IP1, AMFR-V2 and AMFR-V2-C356G-H361A. After 24 h, cells were lysed with 1% Triton X-100 IP buffer (50 mM Tris-HCL, 150 mM NaCl and 0.5 mM EDTA). Lysis buffer without detergents was added to protein lysates to dilute Triton X-100 to 0.3%. Then samples were incubated with HA-agarose beads (Sigma-Aldrich, A2095) or GFP-Trap beads (Chromotek, gta-20) overnight at 4 °C on a rotating platform. Protein-bound beads were washed three times with lysis buffer supplemented with 0.1% Triton X-100 and once with lysis buffer without detergents. HA IP samples were incubated with 40 μl denaturing buffer (2% sodium deoxycholate, 1 mM tris(2-carboxyethyl)phosphine, 4 mM chloroacetamide and 50 mM Tris-HCl pH

8.5) and heated at 95 °C for 10 min. Samples were mixed 1:1 with 500 ng LysC (Promega), incubated for 3 h at 37 °C and digested with 500 ng of trypsin in 50 mM Tris-HCl, pH 8.5, overnight at 37 °C. GFP IP samples were denatured with 25 µl denaturing buffer at 60 °C for 30 min. After cooling down, the samples were digested with 25 µl 50 mM Tris with 1 µl of trypsin (500 ng) at 37 °C overnight. Reactions were stopped by addition of 150 µl of isopropanol containing 1% trifluoroacetic acid (TFA). Peptides were cleaned up by loading them onto SDB-RPS stage tips (Sigma). After one wash with 1% TFA in isopropanol and one wash with 0.2% TFA in water, peptides were eluted using 80% acetonitrile and 1.25% ammonia. Eluted peptides were dried, tandem mass tagged labelled and processed for LC–MS measurements.

LC–MS analysis

Dried peptides were reconstituted in 2% acetonitrile, 0.1% TFA and analysed on a Q Exactive HF mass spectrometer coupled to an easy nLC 1200 (ThermoFisher Scientific) using a 35-cm-long, 75 µm inner diameter fused-silica column packed in-house with 1.9 µm C18 particles (Reprosil pur, Dr. Maisch) and kept at 50 °C using an integrated column oven (Sonation). Peptides were eluted using a nonlinear gradient from 4 to 28% acetonitrile over 45 min and directly sprayed into the mass spectrometer equipped with a nanoFlex ion source (ThermoFisher Scientific). Full scan MS spectra (300–1,650 *m/z*) were acquired in profile mode at a resolution of 60,000 at *m/z* 200, a maximum injection time of 20 ms and an automatic gain control target value of 3×10^6 charges. Up to 15 most intense peptides per full scan were isolated using a 1.4 Th window and fragmented using higher energy collisional dissociation (normalized collision energy of 28). MS/MS spectra were acquired in centroid mode with a resolution of 30,000, a maximum injection time of 45 ms and an automatic gain control target value of 1×10^5. Single charged ions, ions with a charge state above 4 and ions with unassigned charge states were not considered for fragmentation, and dynamic exclusion was set to 20 s to minimize the acquisition of fragment spectra of already acquired precursors.

Proteomics data processing

MS raw data were processed using MaxQuant (v.1.6.10.43) applying default parameters. Acquired spectra were searched against the human 'one sequence per gene' database (taxonomy identifier 9606) downloaded from UniProt (accessed 3 March 2020; 20,531 sequences) and a collection of 244 common contaminants ("contaminants.fasta" provided with MaxQuant) using the Andromeda search engine integrated in MaxQuant[31,32]. Identifications were filtered to obtain false discovery rates (FDRs) below 1% for both peptide spectrum matches (minimum length of 7 amino acids) and proteins using a target–decoy strategy[33]. Protein quantification and data normalization relied on the MaxLFQ algorithm implemented in MaxQuant[34]. The MS proteomics data have been deposited to the ProteomeXchange Consortium[35] through the PRIDE partner repository[36] with the dataset identifiers PXD032718, PXD032720 and PXD039184. All acquired raw files were processed using MaxQuant (v.1.6.10.43) and the implemented Andromeda search engine. For protein assignment, spectra were correlated with the UniProt human database (v.2019) including a list of common contaminants. Searches were performed with tryptic specifications and default settings for mass tolerances for MS and MS/MS spectra. Carbamidomethyl at cysteine residues, oxidations at methionine, acetylation at the N terminus were defined as a fixed modification. The minimal peptide length was set to 7 amino acids and the FDR for proteins and peptide-spectrum matches to 1%. The match-between-run feature with a time window of 1 min was used. For further analysis, Perseus software (v.1.6.6.0) was used and first filtered for contaminants and reverse entries as well as proteins that were only identified by a modified peptide.

Proximity ligation assays

Proximity ligation assays (PLAs) were performed using a Duolink in situ red starter kit mouse/rabbit (DUO92101, Sigma-Aldrich) according to the manufacturer's instructions with rabbit anti-FAM134B (ref. [1]) and rabbit anti-ARL6IP1 (PRS3305, Sigma-Aldrich) antibodies using the Minus (DUO92010) and the Plus probe (DUO92009). Rabbit anti-FAM134B-Plus was used in a 1:5 dilution and rabbit anti-ARL6IP1-Minus in a 1:10 dilution on WT, *Arl6ip1* and *Fam134b* KO MEFs after PFA (4%) fixation and permeabilization with 0.25% (v/v) Triton X-100 in PBS.

Fluorescence protease protection assay

We followed a fluorescence protease protection assay protocol as previously described[37]. In brief, 75,000 COS-7 cells per well were seeded on 18 mm coverslips coated with 0.1 mg ml^{-1} PLL. The next day, cells were transfected with the respective constructs using Lipofectamine 2000. After two more days, cells were washed with pre-warmed intracellular buffer (50 mM HEPES, pH 7.5, 23 mM NaCl, 3 mM $MgCl_2$, 100 nM $CaCl_2$, 1 mM EGTA and freshly added 107 mM potassium glutamate, 1 mM ATP and 2 mM dithiothreitol) and transferred to a heated perfusion chamber filled with the same buffer. Live cell imaging for both GFP and RFP was initiated on a Zeiss Cell Observer Z1 with a frame every 20 s starting with a pre-permeabilization image followed by manual administration of 18 µM digitonin. After 120 s, the buffer was replaced by intracellular buffer containing 6 mM freshly added trypsin. Analysis was carried out using ImageJ by drawing the outline of the selected cell and measuring the mean fluorescence intensity of the surrounded area subtracted by the background intensity taken from a cell free spot of the same frame. For further analysis, the area under the curve was calculated between 160 and 720 s.

Purification of recombinant proteins

Trx–His fusion proteins were purified from *Escherichia coli* treated with 0.5 mM IPTG (overnight, 18 °C). Eluted proteins were concentrated using Amicon Ultra-4-10k centrifugal filter units (Millipore) and then either dialysed at 4 °C against HN buffer (20 mM HEPES/KOH pH 7.4, 150 mM NaCl and 2.5 mM dithiothreitol) (material used for Fig. 4 and Extended Data Fig. 5 analyses) or against liposome buffer (25 mM HEPES-KOH, pH 7.2, 25 mM KCl, 2.5 mM magnesium acetate and 100 mM potassium glutamate) (protein used for studies presented in Fig. 2).

Liposome preparation, liposome incubations, freeze-fracturing and TEM

Liposomes were prepared using Folch-fraction type I lipids (Sigma-Aldrich) according to previously described procedures[10,38]. Liposome co-floatation

assays were performed as previously reported[10]. In brief, liposomes and purified recombinant protein were incubated for 15 min at 37 °C in 0.3 M sucrose in liposome buffer, mixed with 75% sucrose in liposome buffer, overlaid with 200 µl 35% sucrose and 200 µl liposome buffer and then centrifuged at 200,000*g* for 30 min at 28 °C. Six fractions were collected from top to bottom and analysed by SDS–PAGE and fluorescence-based western blotting using a LICOR Odyssey system (LICOR Bioscience).

For shaping assays presented in Fig. 2, 1 mg of liposomes was incubated with 5 µM protein in liposome buffer containing 0.3 M sucrose for 15 min at 37 °C. For shaping assays presented in Fig. 4 and Extended Data Fig. 5, 1 mg of liposomes was incubated at 37 °C with 2.5 µM protein in HN buffer containing 0.3 M sucrose for 15 min. Thereafter, 20 µg proteinase K was added to avoid liposomal aggregates. The reaction was performed for 40 min at 45 °C (ref. [39]). Small aliquots of the liposome suspension were then used for freeze-fracturing. The grids with the samples were systematically explored using an EM 900 electron microscope (Zeiss) operated at 80 kV. Images were acquired with a Wide-angle Dual Speed 2K (Tröndle) CCD camera. The diameters of liposomes were determined using ImageJ.

In vitro ubiquitination assay of ARL6IP1 with recombinant AMFR and sample preparation for MS

For the ubiquitination assay with AMFR, 1 µM purified Trx-His–ARL6IP1, 10 µM ubiquitin (in-house), 10 mM ATP and 10 mM $MgCl_2$ were incubated with 0.8 µM AMFR (provided by B. Schulman, Max Planck Institute of Biochemistry), 100 nM E1 UBA1 (in-house) and 0.8 µM E2 UBE2G2 (Biotechne) in 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, at 37 °C for 2 h. The reaction mixture was analysed by SDS–PAGE and Coomassie staining or immunoblot analysis for GST (Cell Signaling Technology), His (Cell Signaling Technology) and ubiquitin (Cell Signaling Technology) or by mapping of ubiquitinated lysines by MS.

Ubiquitination assays in cells, co-immunoprecipitations and TUBE2 pull-down

The ubiquitination of ARL6IP1 was assessed in HEK293T cells transfected with Myc–ubiquitin, HA–ARL6IP1 and either AMFR–Flag or its catalytically inactive RING mutant. Cells were lysed in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1% Triton X-100, 10 mM *N*-ethylmaleimide and protease inhibitors (Roche Diagnostics, 5892791001)). The lysates were then incubated on ice for 15 min and centrifuged at 12,000*g* at 4 °C for 30 min. Next, 40 µl of the supernatant was collected, mixed with Laemmli sample buffer, boiled for 5 min at 95 °C and stored at –20 °C as input control. Myc-tagged ubiquitinated proteins were immunoprecipitated from lysates cleared with Myc-Trap Agarose (Chromotek, yta-10). Beads were washed three times with lysis buffer, heated at 95 °C for 5 min, subjected to SDS–PAGE and analysed by immunoblotting to detect the HA-Tag. For other immunoprecipitation assays, cleared lysates were incubated with GFP-Trap (Chromotek, gta-20), HA-agarose beads (Sigma-Aldrich, A2095) or TUBE2 agarose beads (Life Sensors, UM402) and incubated at 4 °C overnight. The next day, tubes were centrifuged (800*g*, 4 °C, 5 min), the supernatants removed and the beads washed with ice-cold lysis buffer. Input and co-precipitated fractions were analysed by SDS–PAGE and immunoblotting. For co-immunoprecipitation of endogenous ARL6IP1 and FAM134B, a confluent 15 cm dish of *Arl6ip1* KO or WT MEFs was collected. Lysates were cleared by centrifugation at 12,000*g* for 10 min and incubated with the ARL6IP1 primary antibody at 4 °C overnight. Protein A agarose beads (Roche, 11719408001) were added and incubated at 4 °C for 4 h. Beads were then washed three times with lysis buffer, re-suspended in Laemmli buffer and boiled. Supernatants were analysed by SDS–PAGE and immunoblotting.

Modelling and simulations of ARL6IP1

The atomic model of human ARL6IP1 was built using the AI-based AlphaFold (v.2) program[40]. Five models were constructed, and the top-ranked model was chosen as it had maximal overlap with predicted secondary structures and consensus transmembrane topology, a higher pLDDT score and a relatively lower predicted alignment error (AF confidence measure).

Coarse-grained (CG) molecular dynamics simulations were performed using the MARTINI model (v.2.2)[41,42]. CG models of ubiquitinated and non-ubiquitinated versions were built by using martinize.py[43]. DSSP assignments were used to generate backbone restraints to preserve local secondary structure[44,45]. In the ubiquitinated protein (ARL6IP1-K96-Ub), the iso-peptide bond between K96 and the terminal glycine (G76) of ubiquitin (Protein Data Bank identifier 1UBQ) was modelled by modifying the side chain lysine bead (SC2/+1) into a neutral backbone bead (BB/0) and restraining the distance between the terminal bead of ubiquitin and the lysine side chain to 0.35 nm with a force constant of $k = 1{,}250$ kJ (mol nm^2)$^{-1}$. CG protein models were embedded into POPC (16:0-18:1 PC) bilayers spanning the *xy* plane of a periodic simulation box ($20 \times 20 \times 20$ nm^3) solvated with CG-water containing 150 mM NaCl using the insane.py script[43]. All systems were first energy minimized and then equilibrated using the Berendsen thermostat[46] and barostat[47] along with position restraints on protein backbone beads followed by production runs with a 20 fs time step for a total of 10 µs. The system temperature and pressure were maintained at 310 K and 1 atm with the velocity rescaling thermostat[48] and the semi-isotropic Parrinello–Rahman barostat[49], respectively. All simulations were performed using gromacs (v.2019.3)[50,51].

Statistical analysis

All experiments were performed in at least three independent biological replicates unless indicated otherwise. Data are presented as the mean ± s.e.m. unless indicated otherwise. Data analyses were performed using GraphPad Prism 9. For statistical analysis, raw data were analysed for normal distribution with the Kolmogorov–Smirnov test or with graphical analysis using Q-Q-plot. If appropriate, we used one-way analysis of variance (with Bonferroni post-hoc test unless indicated otherwise), repeated-measures two-way analysis of variance, Kruskal–Wallis *H*-test, Student's *t*-test (two-sided unless indicated otherwise) or Mann–Whitney *U*-test. *P* values less than 0.05 were considered significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MS spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[35] through the PRIDE partner repository[36] with the dataset identifiers PXD032718, PXD032720 and PXD039184. All source data in main and extended data figures are provided as supplementary information. This also includes gels and blots. Materials and associated protocols are available upon request without undue qualifications. Source data are provided with this paper.

References

1. Khaminets, A. et al. Regulation of endoplasmic reticulum turnover by selective autophagy. *Nature* **522**, 354–358 (2015).

2. Kurth, I. et al. Mutations in *FAM134B*, encoding a newly identified Golgi protein, cause severe sensory and autonomic neuropathy. *Nat. Genet.* **41**, 1179–1181 (2009).

3. Novarino, G. et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* **343**, 506–511 (2014).

4. Mochida, K. et al. Receptor-mediated selective autophagy degrades the endoplasmic reticulum and the nucleus. *Nature* **522**, 359–362 (2015).

5. Hübner, C. A. & Kurth, I. Membrane-shaping disorders: a common pathway in axon degeneration. *Brain* **137**, 3109–3121 (2014).

6. Blackstone, C. Converging cellular themes for the hereditary spastic paraplegias. *Curr. Opin. Neurobiol.* **51**, 139–146 (2018).

7. Breuss, M. W. et al. Mutations in *LNPK*, encoding the endoplasmic reticulum junction stabiliser lunapark, cause a recessive

neurodevelopmental syndrome. *Am. J. Hum. Genet.* **103**, 296–304 (2018).

8. Nizon, M. et al. *ARL6IP1* mutation causes congenital insensitivity to pain, acromutilation and spastic paraplegia. *Clin. Genet.* **93**, 169–172 (2018).

9. Chukhrova, A. L. et al. A new case of infantile-onset hereditary spastic paraplegia with complicated phenotype (SPG61) in a consanguineous Russian family. *Eur. J. Neurol.* **26**, e61–e62 (2019).

10. Beetz, C. et al. A spastic paraplegia mouse model reveals REEP1-dependent ER shaping. *J. Clin. Invest.* **123**, 4273–4282 (2013).

11. Zhu, P. P. et al. Transverse endoplasmic reticulum expansion in hereditary spastic paraplegia corticospinal axons. *Hum. Mol. Genet.* **31**, 2779–2795 (2022).

12. Kuroda, M. et al. Determination of topological structure of ARL6ip1 in cells: identification of the essential binding region of ARL6ip1 for conophylline. *FEBS Lett.* **587**, 3656–3660 (2013).

13. Bhaskara, R. M. et al. Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy. *Nat. Commun.* **10**, 2370 (2019).

14. Yamamoto, Y., Yoshida, A., Miyazaki, N., Iwasaki, K. & Sakisaka, T. Arl6IP1 has the ability to shape the mammalian ER membrane in a reticulon-like fashion. *Biochem. J.* **458**, 69–79 (2014).

15. Croucher, D. R. et al. Bimolecular complementation affinity purification (BiCAP) reveals dimer-specific protein interactions for ERBB2 dimers learning the choreography of protein networks. *Sci. Signal.* **9**, ra69 (2016).

16. Smith, M. D. et al. CCPG1 is a non-canonical autophagy cargo receptor essential for ER-phagy and pancreatic ER proteostasis. *Dev. Cell* **44**, 217–232.e11 (2018).

17. González, A. et al. Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum. *Nature* https://doi.org/10.1038/s41586-023-06089-2 (2023).

18. Kumar, D. et al. RTN4B interacting protein FAM134C promotes ER membrane curvature and has a functional role in autophagy. *Mol. Biol. Cell* **32**, 1158–1170 (2021).

19. Shibata, Y. et al. Mechanisms determining the morphology of the peripheral ER. *Cell* **143**, 774–788 (2010).

20. Forrester, A. et al. A selective ER-phagy exerts procollagen quality control via a Calnexin–FAM134B complex. *EMBO J.* https://doi.org/10.15252/embj.201899847 (2019).

21. Wang, S., Tukachinsky, H., Romano, F. B. & Rapoport, T. A. Cooperation of the ER-shaping proteins atlastin, lunapark, and reticulons to generate a tubular membrane network. *eLife* https://doi.org/10.7554/eLife.18605 (2016).

22. Dong, R. et al. The inositol 5-phosphatase INPP5K participates in the fine control of ER organization. *J. Cell Biol.* **217**, 3577–3592 (2018).

23. Fowler, P. C. & O'Sullivan, N. C. ER-shaping proteins are required for ER and mitochondrial network organization in motor neurons. *Hum. Mol. Genet.* **25**, 2827–2837 (2016).

24. Jiang, X. et al. FAM134B oligomerization drives endoplasmic reticulum membrane scission for ER-phagy. *EMBO J.* **39**, e102608 (2020).

25. Fregno, I. et al. ER-to-lysosome-associated degradation of proteasome-resistant ATZ polymers occurs via receptor-mediated vesicular transport. *EMBO J.* **37**, e99259 (2018).

26. Reggio, A. et al. Role of FAM134 paralogues in endoplasmic reticulum remodeling, ER-phagy, and collagen quality control. *EMBO Rep.* **22**, e52289 (2021).

27. Wakil, S. M. et al. Truncating ARL6IP1 variant as the genetic cause of fatal complicated hereditary spastic paraplegia. *BMC Med. Genet.* **20**, 119 (2019).

28. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

29. Verheije, M. H. et al. Mouse hepatitis coronavirus RNA replication depends on GBF1-mediated ARF1 activation. *PLoS Pathog.* **4**, e1000088 (2008).

30. Liang, J. R., Lingeman, E., Ahmed, S. & Corn, J. E. Atlastins remodel the endoplasmic reticulum for selective autophagy. *J. Cell Biol.* **217**, 3354–3367 (2018).

31. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

32. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

33. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).

34. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513–2526 (2014).

35. Deutsch, E. W. et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).

36. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

37. Lorenz, H., Hailey, D. W., Wunder, C. & Lippincott-Schwartz, J. The fluorescence protease protection (FPP) assay to determine protein localization and membrane topology. *Nat. Protoc.* **1**, 276–279 (2006).

38. Wolf, D. et al. Ankyrin repeat-containing N-Ank proteins shape cellular membranes. *Nat. Cell Biol.* **21**, 1191–1205 (2019).

39. Seemann, E. et al. Deciphering caveolar functions by syndapin III KO-mediated impairment of caveolar invagination. *eLife* **6**, e29854 (2017).

40. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

41. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

42. Monticelli, L. et al. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

43. Wassenaar, T. A., Ingolfsson, H. I., Bockmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

44. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

45. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

46. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).

47. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).

48. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

49. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

50. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

51. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).

52. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**, 2101–2102 (2008).

Acknowledgements

We thank staff at the q-Proteomics platform at the Institute of Biochemistry II (Frankfurt) for excellent support and B. Schulman (Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany) for providing recombinant AMFR. This study was funded by the DFG (German Research Foundation) to C.A.H. (RTG 1715, RTG 2155, grants HU 800/6-2, HU 800/13-1 and-2 (FOG 2625) and HU 800/14-1) and I.D. and G.H. (DFG, project number 259130777-SFB 1177). C.A.H. was also supported by the BMBF (TreatHSP 01GM1905D and 01GM2209C). I.D. was also supported by the European Research Council (grant ER-REMODEL), the Else Kroener Fresenius Stiftung, the Dr. Rolf M. Schwiete-Stiftung, and the Max Planck Society. F.R. is supported by ZonMW TOP (91217002), Open Competition ENW-KLEIN (OCENW.KLEIN.118), SNSF Sinergia (CRSII5_189952) and Novo Nordisk Foundation (0066384) grants. M.M. is supported by an ALW Open Programme (ALWOP.355) grant. This work was further supported by DFG grants KE685/7-1 to M.M.K. and QU116/9-1 to B.Q. R.M.B. and G.H.

thank the Center for Supercomputing, GUF and the MPCDF, Garching, for computing resources. G.H. thanks the Max Planck Society for support. The models in Fig. 5a,j were created using BioRender (https://www.biorender.com).

Funding

Open access funding provided by Friedrich Schiller University Jena.

Author information

Author notes

1. Hartmut T. Bocker

   Present address: Blink AG, Jena, Germany

2. Ingo Kurth

   Present address: Institute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University, Aachen, Germany

3. These authors contributed equally: Hector Foronda, Yangxue Fu, Adriana Covarrubias-Pinto, Hartmut T. Bocker

4. These authors jointly supervised this work: Ivan Dikić, Christian A. Hübner

Authors and Affiliations

1. Institute of Human Genetics, Jena University Hospital, Friedrich Schiller University, Jena, Germany

   Hector Foronda, Hartmut T. Bocker, Patricia Franzka, Andrea Bock, Lutz Liebmann, Ingo Kurth & Christian A. Hübner

2. Institute of Biochemistry II, Goethe University School of Medicine, Frankfurt am Main, Germany

Yangxue Fu, Adriana Covarrubias-Pinto, Alexis González, Ramachandra M. Bhaskara, Marina E. Hoffmann & Ivan Dikić

3. Institute for Biochemistry I, Jena University Hospital, Friedrich Schiller University, Jena, Germany

Eric Seemann, Nicole Koch, Michael M. Kessels & Britta Qualmann

4. Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

Ramachandra M. Bhaskara & Ivan Dikić

5. Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt am Main, Germany

Ramachandra M. Bhaskara & Gerhard Hummer

6. Institute of Neuropathology, RWTH Aachen University Hospital, Aachen, Germany

Istvan Katona & Joachim Weis

7. Department of Neurosciences, Rady Children's Institute for Genomic Medicine Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA, USA

Joseph G. Gleeson

8. Department of Biomedical Sciences of Cells and Systems, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Fulvio Reggiori & Muriel Mari

9. Department of Biomedicine, Aarhus University, Aarhus C, Denmark

Fulvio Reggiori & Muriel Mari

10. Aarhus Institute of Advanced Studies (AIAS), Aarhus University, Aarhus C, Denmark

Fulvio Reggiori

11. Institute of Biophysics, Goethe University Frankfurt, Frankfurt am Main, Germany

Gerhard Hummer

12. Center for Rare Diseases, Jena University Hospital, Friedrich Schiller University, Jena, Germany

Christian A. Hübner

Contributions

H.F., Y.F., A.C.-P. and H.T.B. are shared first authors. C.A.H. and I.D. are shared senior authors. H.F., Y.F., A.C.-P., H.T.B., P.F., A.B., L.L. and M.M.: analysis, investigation, methodology and writing. N.K., E.S., I. Katona, A.G. and M.E.H.: analysis, investigation and methodology. R.M.B. and G.H.: modelling and simulations. H.F.: histological analysis of mouse samples, immunofluorescence studies of MEFs and human fibroblasts, autophagy and cell viability assays, immunoblots. Y.F.: immunofluorescence studies of MEFs and different knockdown or knockout cell lines (HEK293T and U2OS), co-immunoprecipitation assays. M.E.H.: MS experiments. A.C.-P. and A.G.: analysis and validation of MS data and ubiquitination assays. H.T.B.: cloning, mouse phenotyping and fluorescence protease protection assay and selected co-immunoprecipitation studies. P.F.: PLAs, analysis of motor end plates and skeletal muscle histology as well as co-immunoprecipitation of ARL6IP1-deletion construct with FAM134B. A.G.: co-immunoprecipitation of FAM134B deletion constructs with ARL6IP1. L.L.: electrophysiology. M.M.: TEM of MEFs and human fibroblasts. A.B.: TEM of DRG somata. I. Katona: TEM of other tissue samples. N.K. and

E.S.: shaping assays. J.G.G.: providing human fibroblasts. F.R.: review and editing. I. Kurth: co-initiation and editing. M.M.K., B.Q. and J.W.: editing and supervision. I.D.: conceptualization, writing, editing and supervision. C.A.H.: initiation, conceptualization, writing, editing and supervision.

Corresponding authors

Correspondence to Ivan Dikić or Christian A. Hübner.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Craig Blackstone and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Phenotype of *Arl6ip1* KO mice.

a) The body weight of KO mice is reduced (2 months: 15 WT and 15 KO mice; 22 months: 9 WT and 11 KO mice; 2-sided unpaired Student's t-test: $p = 0.0034$). **b)** Cortical neurons are decreased in aged KO mice. Cross sections of the motor cortex of 22-month-old WT and KO mice with NeuN-

labelled neurons (1 exp.; 3 mice per genotype; 2-sided unpaired Student's t-test: $p = 0.0312$). Scale bars: 50 µm. **c)** Progressive Purkinje cell loss in KO mice. Representative images of the Purkinje cell layer of HE-stained sagittal cerebellum sections of 22-month-old WT and KO mice (1 exp.; $n = 3$ mice with 3 sections each per genotype; 2-sided unpaired Student's t-test: WT 2 versus WT 22 months, $p = 0.0002$; KO 2 versus KO 22 months $p = 0.0001$, KO versus WT 22 months $p = 0.0001$). Scale bars: 100 µm. **d)** Swollen corticospinal axon in the lumbar spinal cord of a 6-month-old KO mouse (1 exp.). TEM of a horizontal spinal cord section. Scale bar: 5 µm. The higher magnification shows the accumulation of granular electron-dense material, defective organelles and tubulofilamentous material. The axon is surrounded by a distended myelin sheath (arrowheads). Scale bar: 1 µm. **e)** Progressive α-motoneuron loss in the thoracic spinal cord of KO mice. Nissl-stained horizontal sections (1 exp.; 3 mice per genotype; 2-sided unpaired Student's t-test: WT 22 versus KO 22 months $p = 0.0322$, KO 2 versus KO 22 months $p = 0.0018$, WT 2 versus WT 22 months $p = 0.0154$). **f)** The Musculus gastrocnemius mass is reduced in 2-month-old KO mice (1 exp.; 3 WT and 3 KO mice with 2 samples each (left and right); 2-sided unpaired Student's t-test: $p = 0.001$). **g)** Grouped atrophic skeletal muscle fibres in 20-week-old KO mice. Toluidine-blue stained semi-thin sections (1 exp.). Scale bar: 100 µm. **h)** Fragmentation and innervation loss of neuromuscular junctions (NMJ) in the Musculus gastrocnemius of 2-month-old KO mice stained with α-bungarotoxin (red), neurofilament 200 (green) and Hoechst 33258 (1 exp.; 3 mice per genotype). 2-sided unpaired Student's t-test for fragmentation ($p = 0.0012$). **i)** The abundance of some ER proteins with RHDs is changed in brain lysates of 5-month-old KO mice (1 exp. with 5 samples per genotype; 2-sided Mann-Whitney-U-test: Rtn2a+b $p = 0.0317$, Rtn4a $p = 0.0317$, Fam134b $p = 0.0079$, Reep2 $p = 0.0159$). Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 2 ARL6IP1 and FAM134B share structural features and interact.

a) Predicted alignment error of the relative organization of key structural elements of ARL6IP1 shown in Fig. 2b. **b)** Helical wheel representation of

two cytosolic helical segments with characteristics of amphipathic helices with a large hydrophobic moment and net positive charge (blue circle)[52] . **c)** Pairwise sequence alignment of ARL6IP1 and FAM134B indicates preserved RHDs, i.e. two helical hairpins TM1+2 and TM3+4 (grey), and two amphipathic helices, AH_L and AH_C (yellow). Both proteins harbour several predicted ubiquitination sites (red triangles). **d)** Doxycycline induced HA-ARL6IP1 interacts with endogenous FAM134B in U2OS cells (1 exp.). **e)** ARL6IP1 co-precipitates with all known members of the FAM134 family of proteins (1 exp.). The interaction with FAM134B requires the N-terminal part of FAM134B with its first RHD and is independent of its C-terminal coiled-coil domain. Cells were transfected with Myc-ARL6IP1 and the indicated GFP-tagged FAM134B deletion constructs. ATL3, another ER-protein characterised by a RHD, served as a negative control. **f)** The central part of ARL6IP1 with both helical hairpins is involved in the interaction with FAM134B (1 exp.). HEK293T cells were transfected with the indicated HA-tagged ARL6IP1 deletion constructs. WT and variant proteins were pulled down and the endogenous binding partner FAM134B was detected. The deletion sites are indicated as black bars. The replaced lysines for the ARL6IP1-7KR variant are indicated by red bars. Variants marked with * lack the terminal KKNE signal (white bar). **g)** The central part of FAM134B with both helical hairpins is required for the interaction with ARL6IP1 (1exp.). HEK293T cells were transfected with the indicated HA-tagged FAM134B deletion constructs. The deletion sites are indicated as black lines. WT and variant proteins were pulled down and the endogenous binding partner ARL6IP1 detected.

Extended Data Fig. 3 ARL6IP1 forms heterodimers with FAM134B but does not bind LC3B on its own.

a) V1-ARL6IP1/V2-ARL6IP1 homodimers and V1-FAM134B/V2-ARL6IP1 heterodimers are formed in transiently expressing U2OS cells as evidenced by the Venus fluorescence. Co-transfection of V2-ARL6IP1 with V1-CCPG1, another ER resident protein, which does not interact with ARL6IP1, or a construct encoding an ARL6IP1 variant devoid of the first helical hairpin (V2-ARL6IP1Δ(TM1+2)), which is required for the interaction between FAM134B and ARL6IP1, does not result in heteromerisation as evidenced by the lack of the Venus fluorescence. The

single channel images represent immunostaining for either FAM134B, ARL6IP1, and CCPG1 or the Venus signal. 1 experiment. Scale bars: 10 µm. **b)** Venn diagram of interactors of ARL6IP1 and FAM134B homo- and heterodimers, respectively. Numbers represent the identified peptides significantly enriched in three IP and mass spectrometry replicates for each condition. **c,d)** Annotation enrichment analysis of the interactome of V1-ARL6IP1/V2-ARL6IP1 homodimers and V1-FAM134B/V2-ARL6IP1 heterodimers. Bars represent the significantly enriched gene ontology biological process (GOBP), the gene ontology cellular components (GOCC), the gene ontology molecular function (GOMF), and the domain enrichment (Pfam). The Benjamini-Hochberg FDR value is included (right side of the bars). **e)** Western blot analysis showing that RTN4, another ER-shaping protein characterised by the presence of a RHD, and FAM134B co-precipitate with HA-tagged ARL6IP1 (1 exp.).

Extended Data Fig. 4 Mass spectrometry analysis of ubiquitinated lysines in ARL6IP1 and FAM134B proteins.

a,b) Localization of the identified lysines in ARL6IP1 and FAM134B versus Log2 enrichment over mock. Significantly modified lysines (−Log10 >1.3) are highlighted in red. **c–e**) Spectra of modified lysines of ARL6IP1 peptides (K96, K114 and K130) found in ARL6IP1 homo- or heterodimers with FAM134B. **f–l)** Spectra of modified lysines of FAM134B peptides, K90, K160, K374, K278 and K485 found in FAM134B in the heterodimer complex with ARL6IP1.

Extended Data Fig. 5 ER-fragmentation (ER-phagy) is promoted by ubiquitination of ARL6IP1.

a) Co-localisation of either ARL6IP1-HA or ARL6IP1-7KR-HA with Climp63 in *Alr6ip1* KO MEFs (1 exp.; 10 cells per genotype; 2-sided Mann-Whitney-U-test: Pearson's coefficient ARL6IP1-HA versus ARL6IP1-7KR-HA $p = 0.012$, Climp63 area control versus ARL6IP1-HA $p = 0.036$, Climp63 area ARL6IP1-HA versus ARL6IP1-7KR-HA $p = 0.036$). Scale bars: 10 µm. **b)** Co-localisation of either ARL6IP1-HA or ARL6IP1-7KR with FAM134B-Myc in *Alr6ip1* KO MEFs does not differ (1 exp.; 10 cells per genotype; 2-sided Mann-Whitney-U-test). Scale bars: 10 µm. **c)**

Quantitative evaluations of TEM images of freeze-fractured liposomes demonstrate that the shaping properties of His-Trx-ARL6IP1-7KR do not differ from His-Trx-ARL6IP1 (2 exp.; n = 344/391/242; Kruskal-Wallis test with Dunn's post-hoc test: His-Trx-ARL6IP1 versus His-Trx $p < 0.0001$, His-Trx-ARL6IP1-7KR versus His-Trx $p < 0.0001$; n = 344/391/242). **d)** The binding of LC3B to V1-ARL6IP1-7KR/V2-FAM134B heterodimers is reduced compared to V1-ARL6IP1/V2-FAM134B heterodimers (3 exp.). Illustration for quantification shown in Fig. 4f. **e)** The co-localization of ER-fragments with LC3B (red) and Ubiquitin (Ub, white) was analysed in *ARL6IP1* KO U2OS cells transiently expressing either V1-FAM134B/V2-ARL6IP1 or V1-FAM134B/V2-ARL6IP1-7KR heterodimers after treatment with Torin1 for 6 h. Scale bars: 10 µm. Illustration for quantification shown in Fig. 4g. **f)** Endogenous ARL6IP1 is ubiquitinated upon induction of ER stress (1 exp.). U2OS cells were harvested at steady state (basal) or after incubation with the indicated stressors for 6 h (Bafilomycin A1 200 nM, Torin1 250 nM, Thapsigargin 1.5 µM, Tunicamycin 5 µg/ml, Chloroquine 100 µM; stressors and steady state with DMSO 1:1,000). ARL6IP1 is detected in the TUBE2-pull-down. **g)** Upper: The *in vitro* ubiquitination assay with subsequent mass spectrometry shows that AMFR can ubiquitinate ARL6IP1 K96 (1 exp. with 3 replicates; 2-sided unpaired Student's t-test $p = 0.0001$). Lower: Confirmation of ubiquitination of ARL6IP1 by immunoblot analysis (1 exp.). Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 6 FAM134B-driven ER-phagy is compromised in the absence of ARL6IP1.

a) Single channels of the merged images displayed in Fig. 5b and c. Scale bar: 20 µm. **b)** Western blot analysis of siRNA mediated knock-down of ARL6IP1 in HeLa and U2OS cells (1 exp. with 2 replicates). **c)** The ratio between mCherry-positive and mCherry and GFP-positive puncta is decreased in EBSS starved cells upon knock-down of ARL6IP1 and induction of the reporter with doxycycline (1 exp. with triplicates; one-way ANOVA with Bonferroni post-hoc analysis, $p = 0.0001$, $F = 12.19$: siRNA control basal versus EBSS $p = 0.0006$; BafA1 versus EBSS $p = 0.0001$,

EBSS versus EBSS BafA1 p = 0.0001; siRNA control EBSS versus siRNA ARL6IP1 EBSS p = 0.0013). Scale bar: 10 µm. Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 7 ER-phagy but not bulk autophagy, mitophagy or pexophagy is compromised upon disruption of ARL6IP1.

a) siRNA mediated knock-down of either ARL6IP1, FAM134B or both in HeLa cells (3 exp.). **b)** The ratio of Climp63-positive versus Climp63-negative lysosomes upon knock-down of either ARL6IP1 or FAM134B is consistent with a defect in ER-phagy. Cells were treated with Torin1 and Bafilomycin A1. The simultaneous knock-down of both further decreases this ratio (1 exp.; 15 cells per genotype, one-way-ANOVA, F-value = 32.59, p = 0.0001, Bonferroni post-hoc analysis: siCTRL versus siARL6IP1 p = 0.0001, siCTRL versus siFAM134B p = 0.0017, siRNA CTRL versus siARL6IP1+siFAM134B p = 0.0001; siARL6IP1 versus siARL6IP1+siFAM134B p = 0.0001, siFAM134B versus siARL6IP1+siFAM134B p = 0.0001). Scale bar: 5 µm. **c)** Arl6ip1 and Fam134b WT and KO MEFs were EBSS starved for 4 h, fixed and stained for Lamp1, LC3B, and the ER-protein Sec62. The quantification of autophagosomes (LC3B-positive and Lamp1-negative) loaded with ER (Sec62-positive) or devoid of ER (Sec62-negative) supports a defect in ER-phagy but not bulk autophagy (1 exp.; 10 cells per genotype were analysed; one-way-ANOVA with Bonferroni post-hoc analysis; autophagosomes: p = 0.001, F-value = 9.081, WT versus *Arl6ip1* KO p = 0.003, WT versus *Fam134b* KO p = 0.003; autolysosomes: p = 0.004, F-value = 6.954, WT versus *Arl6ip1* KO p = 0.024, WT versus *Fam134b* KO p = 0.005). Scale bar: 20 µm. **d)** Knock-down of ARL6IP1 in U2OS cells with inducible expression of the mCherry-GFP-LC3 reporter (1 exp.). **e)** Bulk autophagy upon siRNA mediated knock-down of ARL6IP1 is not compromised in U2OS cells expressing the mCherry-GFP-LC3 reporter (1 exp. with 3 replicates, one-way-ANOVA with Bonferroni post-hoc analysis). **f)** Mitophagy is not affected by KO of ARL6IP1. WT and *ARL6IP1* KO HeLa cells were transfected with the mitophagy reporter mCherry-GFP-FIS1.

Mitophagy flux was studied at steady state (basal) and after 4 h with 40 µM CCCP (1 exp; n = 11 cells each were analysed; one-way ANOVA with Bonferroni post-hoc analysis, F-value = 13.36, p = 0.0001: WT basal versus CCCP p = 0.0067, KO basal versus CCCP p = 0.0001). Scale bar: 25 µm. **g)** Pexophagy is not affected by ARL6IP1 knock-down. The mCherry-GFP-PMP34 reporter was induced in control and ARL6IP1 knock-down U2OS cells and pexophagy flux studied at steady state (basal) and 20 h of EBSS starvation (1 exp. with n = 9/6/9/8 cells analysed; one-way ANOVA with Bonferroni post-hoc analysis, F-value = 10.65, p = 0.0001: WT basal versus CCCP p = 0.0377, KO basal versus CCCP p = 0.0004). Scale bar: 25 µm. Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 8 Rescue experiment with either ARL6IP1 or the ARL6IP1-7KR variant.

a) In *Arl6ip1* KO MEFs ER-phagy can be rescued by overexpression of either ARL6IP1 or ARL6IP1-7KR-HA. KO MEFs were transfected with the mCherry-GFP-FAM134B reporter and either ARL6IP1-HA or ARL6IP1-7KR-HA, fixed and stained for LC3B. Quantifications of LC3B/mCherry/GFP-positive puncta (autophagosomes) and LC3B/mCherry-positive but GFP-negative puncta (autolysosomes) per cell area (1 exp. with 15 cells analysed per genotype; Kruskal-Wallis test with Dunn`s multiple comparison test: autophagosomes WT versus KO p = 0.0002, KO versus ARL6IP1 rescue p = 0.0001, KO versus ARL6IP1-7KR rescue p = 0.0138; autolysosomes WT versus KO p = 0.001, KO versus ARL6IP1 rescue p = 0.0001, KO versus ARL6IP1-7KR rescue p = 0.0053). Scale bars: 20 µm. Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 9 Intracellular accumulation of collagen in the absence of ARL6IP1.

a) The ratio of cells strongly labelling for collagen is increased in both *Fam134b* and *Arl6ip1* KO MEFs (1 exp. with n = 86 WT cells, n = 121

FAM134B KO MEF cells, n = 69 ARL6IP1 KO MEFs analysed; one-way ANOVA with Bonferroni post-hoc analysis, F-value = 171.7, p = 0.0001: WT versus *Fam134b* KO p = 0.0001, WT versus *Arl6ip1* KO p = 0.0001). Scale bar: 50 µm. **b)** Comparable collagen levels in the supernatants of *Fam134b* and *Arl6ip1* KO MEFs exclude a defect of collagen secretion (1 exp. with n = 2 samples). **c)** The ratio of cells strongly labelling for collagen is increased in fibroblasts of the patient homozygous for the K193Ffs variant (n = 3 experiments with 344 (control), 310 (K193Ffs het.), 399 (K193Ffs hom.) cells analysed; one-way ANOVA with Bonferroni post-hoc analysis, F = 40.02, p = 0.0001: WT versus K193Ffs hom. p = 0.0001, K193Ffs het. versus K193Ffs hom. p = 0.0001). **d)** Comparable collagen levels in the supernatants of fibroblasts of a healthy control, the heterozygous father and the affected patient exclude a defect of collagen secretion (1 exp. with n = 2 samples). **e)** Western blot analysis of WT, *Fam134b* KO and *Arl6ip1* KO MEFs at steady state and EBSS starvation with or without Bafilomycin A1 (BafA). The accumulation of collagen is not enhanced upon Bafilomycin A1 treatment. This suggests that the removal of misfolded collagen by ER-phagy is abolished in *Arl6ip1* KO MEFs. Vinculin served as loading control (1 exp.). **f)** Accumulation of collagen in patient fibroblasts. Actin served as loading control (3 experiments; one-way ANOVA with Bonferroni post-hoc analysis, F-value = 40.8, p = 0.0003: WT versus K193Ffs hom. p = 0.0011; K193Ffs het. versus K193Ffs hom. p = 0.0005). **g)** MEFs devoid of Fam134b or Arl6ip1 cannot deliver collagen to lysosomes. Quantification of collagen-positive lysosomes per cell in cells loaded with collagen (1 exp., 6 cells analysed with 564 (WT), 474 (*Fam134b* KO), 785 (*Arl6ip1* KO) puncta counted; one-way ANOVA with Bonferroni post-hoc analysis, F-value 47.59, p = 0.0001: WT versus *Fam134b* KO p = 0.0001, WT versus *Arl6ip1* KO p = 0.0001, *Fam134b* KO versus *Arl6ip1* KO p = 0.0165). Scale bars: 20 µm. Quantitative data are shown as mean ± SEM.

Source Data

Extended Data Fig. 10 Consequences of compromised FAM134B-dependent ER-phagy for ER structure and cell viability.

a) Staining for the ER sheet protein Climp63 and the ER tubule marker RTN4 in WT and *Fam134b* and *Arl6ip1* KO MEFs. Relative Climp63-positive or RTN4-positive pixels per cell and their ratio was calculated (3 exp. with 45 cells analysed per genotype; Kruskal-Wallis test with Dunn's multiple comparisons test: Climp63 WT versus *Arl6ip1* KO p = 0.0001, WT versus *Fam134b* KO p = 0.0001; RTN4 WT versus *Arl6ip1* KO p = 0.036, WT versus *Fam134b* KO p = 0.0005; data points obtained within the same experiment are indicated by the same colour). Scale bar: 10 µm. **b)** Compared with the strongly increased number of small highly curved ER protrusions emanating from ER sheets in fibroblasts of the homozygous patient, fibroblasts of the heterozygous father show an intermediate phenotype (1 exp. with 55 cells per genotype; one-way ANOVA with Bonferroni post-hoc analysis, F-value 42,59, p = 0.0001: WT versus K193Ffs het. p = 0.0002, WT versus K193Ffs hom. p = 0.0001, K193Ffs hom. versus K193Ffs het. p = 0.0001; data points obtained within the same experiment are indicated by the same colour). Scale bar: 500 nm. The ER is shown in light purple colour. ER-emerging spikes are shown in pink. **c)** In the presence of the ER stressor Tunicamycin without the proteasomal inhibitor MG132 the viability of the fibroblasts of the heterozygous father is slightly increased compared to control (3 experiments with 2 replicates each; 2-sided Mann-Whitney-U-test; Thapsigargin: WT versus K193Ffs hom. p = 0.0022, K193Ffs het. versus K193Ffs hom. p = 0.0022; Thapsigargin + MG132: WT versus K193Ffs hom. p = 0.0087; Tunicamycin: WT versus K193Ffs het. p = 0.0411, WT versus K193Ffs hom. p = 0.0022; K193Ffs het. versus K193Ffs hom. p = 0.0022; Tunicamycin+MG132: WT versus K193Ffs hom. p = 0.026). Individual experiments are indicated by differently coloured data points. Quantitative data are shown as mean ± SEM.

Source Data

Supplementary information

Supplementary Information

Western blots.

Reporting Summary

Supplementary Table 1

Plasmids used in the study.

Supplementary Table 2

Primers used for this study.

Supplementary Table 3

Primary antibodies and their application related to this study.

Supplementary Table 4

Secondary antibodies and their application related to this study.

Supplementary Table 5

sgRNA sequences used to generate the *ARL6IP1* KO cell lines.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Source Data Fig. 5

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Foronda, H., Fu, Y., Covarrubias-Pinto, A. *et al.* Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy. *Nature* **618**, 402–410 (2023). https://doi.org/10.1038/s41586-023-06090-9

- Received: 19 March 2022

- Accepted: 17 April 2023
- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06090-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum**
    - Alexis González
    - Adriana Covarrubias-Pinto
    - Ivan Dikić

  Nature (2023)

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06090-9

Amendments & Corrections

- **Author Correction: Magnetic memory and spontaneous vortices in a van der Waals superconductor** [24 May 2023]
- **Author Correction: Coherent correlation imaging for resolving fluctuating states of matter** [24 May 2023]
- **Author Correction: Dopamine promotes head direction plasticity during orienting movements** [24 May 2023]
- **Publisher Correction: The NK cell receptor NKp46 recognizes ecto-calreticulin on ER-stressed cells** [17 May 2023]
- **Addendum: Systems vaccinology of the BNT162b2 mRNA vaccine in humans** [24 May 2023]

- Author Correction
- Published: 24 May 2023

Author Correction: Magnetic memory and spontaneous vortices in a van der Waals superconductor

- Eylon Persky ,
- Anders V. Bjørlig,
- Irena Feldman,
- Avior Almoalem,
- Ehud Altman,
- Erez Berg,
- Itamar Kimchi,
- Jonathan Ruhman,
- Amit Kanigel &
- …
- Beena Kalisky

Show authors

Nature **volume 618**, page E14 (2023)

- 322 Accesses

- 1 Altmetric

- Metrics details

The Original Article was published on 27 July 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-04855-2
Published online 27 July 2022

In the version of this article initially published, the images presented in Extended Data Figure 2 were inadvertent duplicates of Figure 2c–e and are now updated in the HTML and PDF versions of the article.

Author information

Authors and Affiliations

1. Department of Physics, Bar Ilan University, Ramat Gan, Israel

   Eylon Persky, Anders V. Bjørlig, Jonathan Ruhman & Beena Kalisky

2. Bar Ilan Institute of Nanotechnology and Advanced Materials, Bar Ilan University, Ramat Gan, Israel

   Eylon Persky, Anders V. Bjørlig & Beena Kalisky

3. Department of Physics, Technion–Israel Institute of Technology, Haifa, Israel

   Irena Feldman, Avior Almoalem & Amit Kanigel

4. Department of Physics, University of California, Berkeley, Berkeley, CA, USA

   Ehud Altman

5. Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

   Ehud Altman

6. Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot, Israel

Erez Berg

7. School of Physics, Georgia Institute of Technology, Atlanta, GA, USA

Itamar Kimchi

Corresponding authors

Correspondence to Eylon Persky or Beena Kalisky.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Persky, E., Bjørlig, A.V., Feldman, I. *et al.* Author Correction: Magnetic memory and spontaneous vortices in a van der Waals superconductor. *Nature* **618**, E14 (2023). https://doi.org/10.1038/s41586-023-06196-0

- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06196-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06196-0

- Author Correction
- Open Access
- Published: 24 May 2023

Author Correction: Coherent correlation imaging for resolving fluctuating states of matter

- Christopher Klose,
- Felix Büttner ,
- Wen Hu ,
- Claudio Mazzoli,
- Kai Litzius,
- Riccardo Battistelli,
- Sergey Zayko,
- Ivan Lemesh,
- Jason M. Bartell,
- Mantao Huang,
- Christian M. Günther,
- Michael Schneider,
- Andi Barbour,
- Stuart B. Wilkins,
- Geoffrey S. D. Beach,
- Stefan Eisebitt &
- …
- Bastian Pfau

Show authors

Nature **volume 618**, page E15 (2023)

- 250 Accesses

- 1 Altmetric

- Metrics details

The Original Article was published on 18 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05537-9
Published online 18 January 2023

In the version of this article originally published, Sergey Zayko was missing from the author list and has now been included in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Claudio Mazzoli, Bastian Pfau

Authors and Affiliations

1. Max Born Institute, Berlin, Germany

   Christopher Klose, Michael Schneider, Stefan Eisebitt & Bastian Pfau

2. Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

   Felix Büttner, Kai Litzius, Ivan Lemesh, Jason M. Bartell, Mantao Huang & Geoffrey S. D. Beach

3. National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY, USA

   Felix Büttner, Wen Hu, Claudio Mazzoli, Andi Barbour & Stuart B. Wilkins

4. Helmholtz-Zentrum für Materialien und Energie, Berlin, Germany

   Felix Büttner & Riccardo Battistelli

5. IV Physical Institute, University of Göttingen, Göttingen, Germany

   Sergey Zayko

6. Zentraleinrichtung Elektronenmikroskopie (ZELMI), Technische Universität Berlin, Berlin, Germany

   Christian M. Günther

7. Institut für Optik und Atomare Physik, Technische Universität Berlin, Berlin, Germany

   Stefan Eisebitt

Corresponding authors

Correspondence to Felix Büttner or Wen Hu.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Klose, C., Büttner, F., Hu, W. *et al.* Author Correction: Coherent correlation imaging for resolving fluctuating states of matter. *Nature* **618**, E15 (2023). https://doi.org/10.1038/s41586-023-06224-z

- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06224-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06224-z

- Author Correction
- Open Access
- Published: 24 May 2023

Author Correction: Dopamine promotes head direction plasticity during orienting movements

- Yvette E. Fisher,
- Michael Marquis,
- Isabel D'Alessandro &
- …
- Rachel I. Wilson

Show authors

Nature **volume 618**, page E16 (2023)

- 342 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 30 November 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05485-4
Published online 30 November 2022

In the version of this article initially published, the deposit listed in the Code availability section did not include licence information for previously

published code; the section is now updated to read “Custom code is available at https://doi.org/10.5281/zenodo.7930823” in the HTML and PDF versions of the article.

Author information

Author notes

1. Yvette E. Fisher

   Present address: Department of Molecular and Cellular Biology, University of California Berkeley, Berkeley, CA, USA

2. Yvette E. Fisher

   Present address: Chan Zuckerberg Biohub, San Francisco, CA, USA

3. These authors contributed equally: Yvette E. Fisher, Michael Marquis

Authors and Affiliations

1. Department of Neurobiology, Harvard Medical School, Boston, MA, USA

   Yvette E. Fisher, Michael Marquis, Isabel D’Alessandro & Rachel I. Wilson

Corresponding author

Correspondence to Rachel I. Wilson.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Fisher, Y.E., Marquis, M., D'Alessandro, I. *et al.* Author Correction: Dopamine promotes head direction plasticity during orienting movements. *Nature* **618**, E16 (2023). https://doi.org/10.1038/s41586-023-06238-7

- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06238-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06238-7

- Publisher Correction
- Published: 17 May 2023

Publisher Correction: The NK cell receptor NKp46 recognizes ecto-calreticulin on ER-stressed cells

- Sumit Sen Santara,
- Dian-Jang Lee,
- Ângela Crespo,
- Jun Jacob Hu,
- Caitlin Walker,
- Xiyu Ma,
- Ying Zhang,
- Sourav Chowdhury,
- Karla F. Meza-Sosa,
- Mercedes Lewandrowski,
- Haiwei Zhang,
- Marjorie Rowe,
- Arthur McClelland,
- Hao Wu,
- Caroline Junqueira &
- …
- Judy Lieberman

Show authors

Nature **volume 618**, page E17 (2023)

- 510 Accesses

- 2 Altmetric

- Metrics details

The Original Article was published on 05 April 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-023-05912-0
Published online 5 April 2023

In the version of this article initially published, author edits were missing and have now been incorporated. In the Fig. 2f legend now reading “CRT-Fc proteins … and immunoblot was probed with anti-Fc (right),” the direction has been added; in Fig. 3c, the bottom-right *x*-axis label reading “OP” has been restored; panel citations in the first sentence of the Fig. 4 legend were clarified to indicate “WT (**a**,**b**) and *Ncr1*$^{-/-}$ mice (**a**–**c**)”; in Extended Data Fig. 3d, the residue on the bottom left reading “K170” has been corrected from “K17”; the second sentence of the Extended Data Fig. 5 caption has been added, now reading “In **a**, 2 clones were analysed (CL1 and CL2), but **b**–**f** show data for only clone 2 (CL2)”; the number of cell lines in the Extended Data Fig. 8c caption has been adjusted from 2 to read “n = 3 cell line samples” and the end of the Extended Data Fig. 8 legend has been amended to read “unpaired two-way ANOVA (**b**–**d**)” in place of “**a**–**d**”. The changes have been made in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Sumit Sen Santara, Dian-Jang Lee

Authors and Affiliations

1. Program in Cellular and Molecular Medicine, Boston Children’s Hospital, Boston, MA, USA

Sumit Sen Santara, Dian-Jang Lee, Ângela Crespo, Jun Jacob Hu, Caitlin Walker, Xiyu Ma, Ying Zhang, Karla F. Meza-Sosa, Mercedes Lewandrowski, Haiwei Zhang, Marjorie Rowe, Hao Wu, Caroline Junqueira & Judy Lieberman

2. Department of Pediatrics, Harvard Medical School, Boston, MA, USA

   Sumit Sen Santara, Dian-Jang Lee, Ângela Crespo, Xiyu Ma, Ying Zhang, Karla F. Meza-Sosa, Mercedes Lewandrowski, Haiwei Zhang, Marjorie Rowe, Caroline Junqueira & Judy Lieberman

3. Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, India

   Sumit Sen Santara

4. Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA

   Jun Jacob Hu, Caitlin Walker & Hao Wu

5. Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

   Sourav Chowdhury

6. Laboratorio de Neuroinmunobiología, Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

   Karla F. Meza-Sosa

7. Center for Nanoscale Systems, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA

   Arthur McClelland

8. Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil

Caroline Junqueira

Corresponding authors

Correspondence to Caroline Junqueira or Judy Lieberman.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Sen Santara, S., Lee, DJ., Crespo, Â. *et al.* Publisher Correction: The NK cell receptor NKp46 recognizes ecto-calreticulin on ER-stressed cells. *Nature* **618**, E17 (2023). https://doi.org/10.1038/s41586-023-06151-z

- Published: 17 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06151-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06151-z

- Addendum
- Published: 24 May 2023

Addendum: Systems vaccinology of the BNT162b2 mRNA vaccine in humans

- Prabhu S. Arunachalam,
- Madeleine K. D. Scott,
- Thomas Hagan,
- Chunfeng Li,
- Yupeng Feng,
- Florian Wimmers,
- Lilit Grigoryan,
- Meera Trisal,
- Venkata Viswanadh Edara,
- Lilin Lai,
- Sarah Esther Chang,
- Allan Feng,
- Shaurya Dhingra,
- Mihir Shah,
- Alexandra S. Lee,
- Sharon Chinthrajah,
- Sayantani B. Sindher,
- Vamsee Mallajosyula,
- Fei Gao,
- Natalia Sigal,
- Sangeeta Kowli,
- Sheena Gupta,
- Kathryn Pellegrini,
- Gregory Tharp,
- Sofia Maysel-Auslender,

- Sydney Hamilton,
- Hadj Aoued,
- Kevin Hrusovsky,
- Mark Roskey,
- Steven E. Bosinger,
- Holden T. Maecker,
- Scott D. Boyd,
- Mark M. Davis,
- Paul J. Utz,
- Mehul S. Suthar,
- Purvesh Khatri ,
- Kari C. Nadeau &
- …
- Bali Pulendran

Show authors

Nature **volume 618**, page E18 (2023)

- 576 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 12 July 2021

Addendum to: *Nature* https://doi.org/10.1038/s41586-021-03791-x
Published online 12 July 2021

Since the publication of our original article, we have been alerted to a discrepancy between the expression of sex-specific genes in the cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) and bulk transcriptomics datasets and the reported gender of the subjects, whose transcriptomes were analysed in these datasets.

CITE-seq: The reported gender of the six participants included in the CITE-seq analysis was correct. However, there was a cross-contamination of the day-1 samples obtained from participants 2049 (male) and 2051 (female), which resulted in the erroneous appearance of cell clusters containing cells of both male and female origin in the CITE-seq data. By analysing single-nucleotide variants, we correctly reassigned the contaminating cells to the appropriate subjects and reanalysed the CITE-seq data. This reanalysis resulted in 242,202 high-quality transcriptomes, as opposed to 242,479 transcriptomes reported previously and resulted in minor changes to Fig. 4c and i and Extended Fig. 10f, but did not affect any of the conclusions.

Bulk transcriptomics: We also discovered that an inadvertent mislabelling of samples had occurred at the genomics core facility during the processing of samples for bulk transcriptomics. Based on the original lab records, we have now reassigned the correct samples to the appropriate subjects and time points. We independently confirmed our reassignment using single-nucleotide polymorphism analysis of banked PBMC samples from each corresponding subject. The correct reassignment resulted in changes to several figures (Figs. 3 and 5 and Extended Data Figs. 1a, 8, 11, and 12) and Supplementary Table 1. The major conclusions from the analysis of bulk transcriptional data, including the dynamics and nature of the transcriptional response, and the transcriptional correlates of T cell and antibody response remain unaffected. Furthermore, the central observation of an enhanced innate transcriptional signature following secondary vaccination was strengthened (Fig. 3) and has now been confirmed by independent studies[1,2]. However, the reassignment of samples to the appropriate time points resulted in more negatively enriched gene sets on day 7 and later, and a diminution of the correlation between the monocyte-associated transcriptional signature and the cross-neutralization potential against the B.1.351 strain (Extended Data Fig. 12). The negatively enriched modules represent innate pathways whose magnitudes at these time points (i.e., days 7, 21, and 28) were negligible relative to their magnitude at day 22 (peak of the response). As such, we do not make strong biological conclusions based on this negligible expression of the innate immunity modules on days 7, 21, and 28. The revised figures and description of the results are incorporated in the online version of the paper.

1. 1.

   On page 411, in the section “Transcriptional signatures of vaccination,” “31 participants” now reads “32 participants”, “Six of 185” is now “Four of 185”, and “fourfold” is now “twelvefold” in the first three sentences of the paragraph.

2. 2.

   On page 413, in the section “Comparison with other vaccines,” the third sentence now begins with “The responses at day 1 after the prime and boost doses of BNT162b2 were broadly similar”, instead of the original “Although the day-1 response to the first dose of BNT162b2 showed little overlap with that of other vaccines, the response at day 1 after the boost was broadly similar”. Furthermore, in the sentence beginning, “Meanwhile, day-7 responses”, which carries over to page 414, the sentence now ends with “but cell-cycle-related transcriptional modules after both doses were shared with many vaccines”, rather than the original “with cell-cycle-related transcriptional modules after the prime dose being the signature shared with many vaccines”.

3. 3.

   On page 415, in the fifth sentence of the section “Transcriptional correlates of adaptive immunity,” “monocyte and inflammatory modules were highly associated” has been replaced with “a monocyte module was associated”.

The corrected and the published version of the figures and the details of the sample assignment are presented in Supplementary Figs. 1–7 and Supplementary Tables 1 and 2, respectively, for transparency.

Supplementary information is available in the online version of this Amendment.

References

1. Yamaguchi, Y. et al. Consecutive BNT162b2 mRNA vaccination induces short-term epigenetic memory in innate immune cells. *JCI Insight* **7**, e163347 (2022).

2. Rinchai, D. et al. High-temporal resolution profiling reveals distinct immune trajectories following the first and second doses of COVID-19 mRNA vaccines. *Sci. Adv.* **8**, eabp9961 (2022).

Author information

Author notes

1. These authors contributed equally: Prabhu S. Arunachalam, Madeleine K. D. Scott, Thomas Hagan

Authors and Affiliations

1. Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA

   Prabhu S. Arunachalam, Madeleine K. D. Scott, Chunfeng Li, Yupeng Feng, Florian Wimmers, Lilit Grigoryan, Meera Trisal, Sarah Esther Chang, Allan Feng, Shaurya Dhingra, Vamsee Mallajosyula, Fei Gao, Natalia Sigal, Sangeeta Kowli, Sheena Gupta, Sofia Maysel-Auslender, Holden T. Maecker, Mark M. Davis, Paul J. Utz, Purvesh Khatri & Bali Pulendran

2. Center for Biomedical Informatics, Department of Medicine, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

   Madeleine K. D. Scott & Purvesh Khatri

3. Division of Infectious Diseases, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA

   Thomas Hagan

4. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Thomas Hagan & Steven E. Bosinger

5. Yerkes National Primate Research Center, Atlanta, GA, USA

Venkata Viswanadh Edara, Lilin Lai, Kathryn Pellegrini, Gregory Tharp, Sydney Hamilton, Hadj Aoued & Mehul S. Suthar

6. Department of Medicine, Division of Immunology and Rheumatology, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

Sarah Esther Chang, Allan Feng, Shaurya Dhingra & Paul J. Utz

7. Sean N. Parker Center for Allergy and Asthma Research, Stanford University, Stanford, CA, USA

Mihir Shah, Alexandra S. Lee, Sharon Chinthrajah, Sayantani B. Sindher, Scott D. Boyd & Kari C. Nadeau

8. Quanterix, Billerica, MA, USA

Kevin Hrusovsky & Mark Roskey

9. Department of Pathology, Emory University School of Medicine, Atlanta, GA, USA

Steven E. Bosinger

10. Department of Pathology, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

Scott D. Boyd, Mark M. Davis & Bali Pulendran

11. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

Mark M. Davis & Bali Pulendran

12. Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

Kari C. Nadeau

13. Department of Medicine, Division of Pulmonary, Allergy and Critical Care Medicine, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

Kari C. Nadeau

Corresponding authors

Correspondence to Purvesh Khatri, Kari C. Nadeau or Bali Pulendran.

Supplementary information

Supplementary Figures

Original and revised Figs. 3, 4c, 4i, and 5 and Extended Data Figs. 1a, 8, 10f, 11, and 12.

Supplementary Table 1

.

Supplementary Table 2

.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Arunachalam, P.S., Scott, M.K.D., Hagan, T. *et al.* Addendum: Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature* **618**, E18 (2023). https://doi.org/10.1038/s41586-023-05977-x

- Published: 24 May 2023
- Issue Date: 08 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05977-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05977-x

Nature Outlook

- **Smoking** [07 Jun 2023]
  • Tobacco smoking places an enormous burden on global health — around 14% of all deaths in 2019 were attributed to this habit.

| Main menu | Previous section |

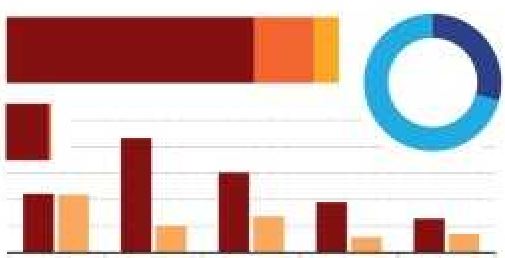


The battle against tobacco is not yet won

Efforts to discourage new smokers and help people to quit are bringing smoking rates down in many places, but it’s going to take more to put an end to tobacco smoking altogether.

- Richard Hodson

Outlook7 Jun 2023 Nature



Smoking: an avoidable health disaster explained

Statistics show that, despite some progress towards smoking cessation, tobacco continues to place an extraordinarily heavy burden on global health.

- Richard Hodson

Outlook7 Jun 2023 Nature



African countries fight for tobacco control

A lack of resources and interference from the tobacco industry make it difficult to strengthen anti-smoking regulations.

- Tammy Worth

Outlook7 Jun 2023 Nature



Brain-zapping technology helps smokers to quit

Repetitive transcranial magnetic stimulation is already approved to help people overcome addiction to cigarettes, but researchers still have a lot to learn about how to deliver the treatment effectively.

- Simon Makin

Outlook7 Jun 2023 Nature



[Is nicotine bad for long-term health? Scientists aren’t sure yet]

As e-cigarettes grow in popularity, the lack of knowledge about whether nicotine contributes to the damaging health effects of smoking is becoming more worrying.

- Anthony King

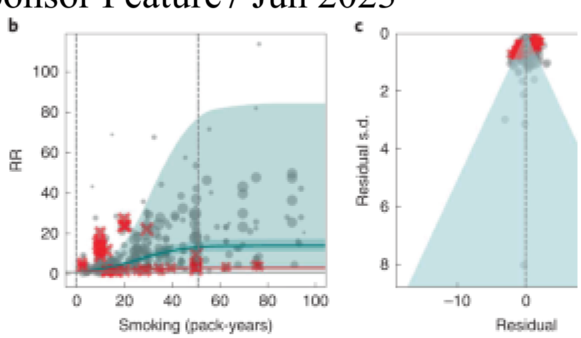
Outlook7 Jun 2023 Nature



[Breaking nicotine’s chokehold]

Established nicotine replacement therapies combined with behavioural counselling could offer the best chance of beating nicotine addiction.

Sponsor Feature7 Jun 2023



Health effects associated with smoking: a Burden of Proof study

A meta-analysis using the Burden of proof method reported consistent evidence supporting harmful associations between smoking and 28 different health outcomes.

- Xiaochen Dai
- Gabriela F. Gil
- Emmanuela Gakidou

ArticleOpen Access10 Oct 2022 Nature Medicine

Every child deserves a smoke-free home

- Lindy U. McGee
- Jean L. Raphael
- Mona Patel

Comment14 Oct 2022 Pediatric Research

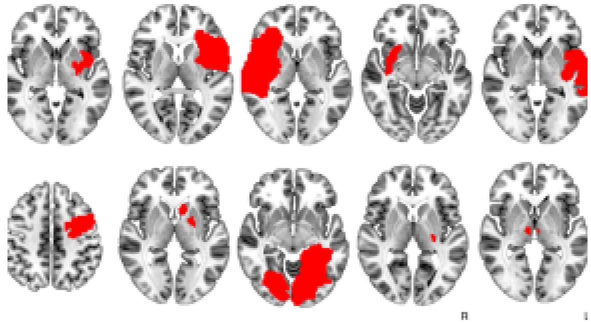


Nicotine e-cigarettes as a tool for smoking cessation

There is abundant evidence that e-cigarettes can help some individuals to quit smoking, so they should be more widely recommended as smoking cessation aids.

- Kenneth E. Warner
- Neal L. Benowitz
- Nancy A. Rigotti

Comment13 Feb 2023 Nature Medicine

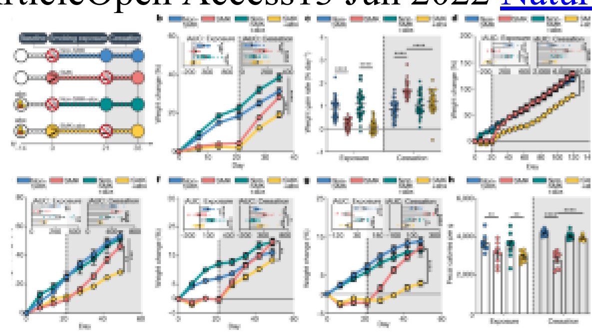


Brain lesions disrupting addiction map to a common human brain circuit

Lesions resulting in addiction remission occur in multiple different brain locations but map to a specific brain circuit.

- Juho Joutsa
- Khaled Moussawi
- Michael D. Fox

ArticleOpen Access13 Jun 2022 Nature Medicine



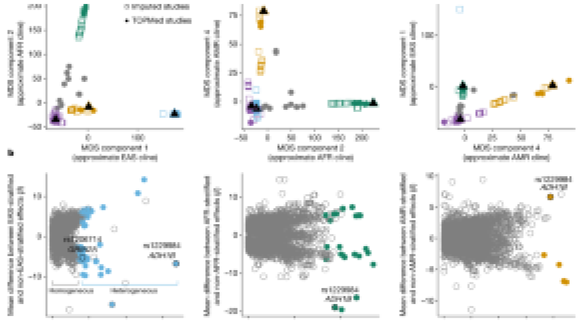
Gut microbiota modulates weight gain in mice after discontinued smoke exposure

A study of mice exposed to cigarette smoke suggests that smoking-cessation-induced weight gain is associated with a dysbiotic state that is driven by smoking-related metabolites.

- Leviel Fluhr
- Uria Mor

- Eran Elinav

Article8 Dec 2021 [Nature]

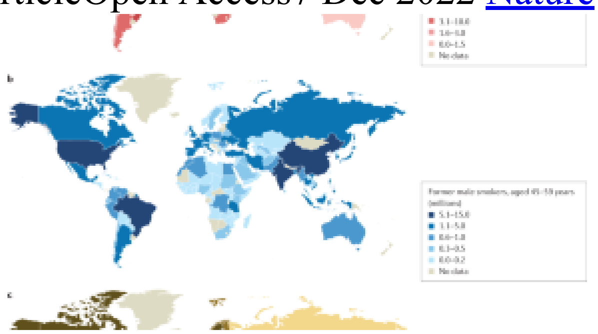


[Genetic diversity fuels gene discovery for tobacco and alcohol use]

A multi-ancestry meta-regression study analyses diverse genome-wide association studies and genome loci associated with tobacco and alcohol use.

- Gretchen R. B. Saunders
- Xingyan Wang
- Scott Vrieze

ArticleOpen Access7 Dec 2022 [Nature]

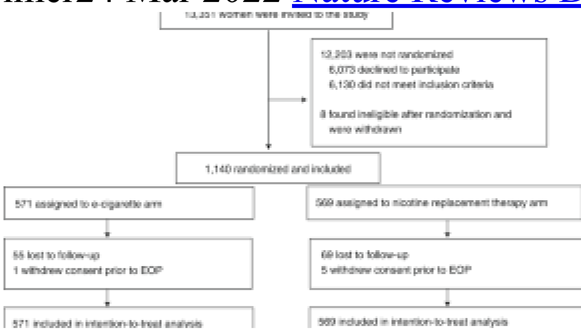


[Tobacco and nicotine use]

This Primer describes the epidemiology, mechanisms, diagnosis and management of tobacco use and tobacco use disorder. Moreover, this Primer discusses the quality of life issues associated with tobacco use and provides an overview of future research avenues for this field.

- Bernard Le Foll
- Megan E. Piper
- Wayne D. Hall

Primer24 Mar 2022 Nature Reviews Disease Primers



Electronic cigarettes versus nicotine patches for smoking cessation in pregnancy: a randomized controlled trial

A randomized controlled trial comparing the efficacy of electronic cigarettes and nicotine patches for smoking cessation in pregnant women found no differences between the interventions. However, electronic cigarettes were found to have some benefit when only the data from women who adhered strictly to the trial protocol were analyzed.

- Peter Hajek
- Dunja Przulj
- Katie Myers Smith

ArticleOpen Access16 May 2022 Nature Medicine

This article was downloaded by **calibre** from https://www.nature.com/collections/dhfebbibdj